

```

QY      1 SSGPSL 6
Db      155 SSGPSL 160

RESULT 6
Q9D8Y9 PRELIMINARY; PRT; 214 AA.
AC Q9D8Y9;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE 1810018L0SRik protein.
GN 1810018L0SRik.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Pancreas;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA Saizawa T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Mateuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staib F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bona M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gusninch S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Marzarelli J., Nombarts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyo-oka K., Wang K.-H., Weitz C., Whitaker C., Wilming L.,
RA Hayashizaki Y.,
RA Hynes-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohsaki S.,
RT "functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
DR EMBL; AK007533; BAB25094.1;
DR MGD; MG1:1917630; 1810018L0SRik.
DR GO; GO:0005198; F:structural molecule activity; IEA.
DR InterPro; IPR000535; MSP domain.
DR Pfam; PF00635; MSP domain; 1.
DR PROSITE; PS50202; MSP; 1.
SQ SEQUENCE 214 AA; 24130 MW; ECD4417E8933CF01 CRC64;

Query Match 100.0%; Score 29; DB 11; Length 214;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 SSGPSL 6
Db      156 SSGPSL 161

RESULT 7
Q7SYQ6 PRELIMINARY; PRT; 215 AA.
AC Q7SYQ6;
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein.
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae;
OC Xenopodinae; Xenopus.
OX NCBI_TaxID=8355;

```

```

RN      SEQUENCE FROM N.A.
RP TISSUE=Whole;
RX MEDLINE=22341132; PubMed=12454917;
RA Klein S.L., Strausberg R.L., Wagner L., Pontius J., Clifton S.W.,
RA Richardson P.;
RT "Genetic and genomic tools for Xenopus research: The NIH Xenopus
RL Dev. Dyn. 225:384-391(2002).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Whole;
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins S.P., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.F., Casavant T.L., Scheetz T.E.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Carninci P., Prange C.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Abramson R.D., Mullahy S.J.,
RA Raha S.S., Loquellano N.A., Peters G.J., Malek J.A., Gunaratne P.H.,
RA Rosak S.A., McEwan P.J., McKernan K.J., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Richards S., Worley K.C., Hale S., Sodergren E.J., Lu X., Gibbs R.A.,
RA Villalón D.K., Musny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahy J., Helton E., Kettman M., Madan A.C., Shevchenko Y., Bouffard G.G.,
RA Whiting M., Madan A., Young A.C., Green E.D., Dickson M.C.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzywicki M.I., Skaleka U., Smailus D.E., Schnerch A., Schein J.E.,
RA Jones S.J., Maria M.A.,
RT "generation and initial analysis of more than 15,000 full-length human
RL and mouse cDNA sequences.";
RN Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [3]
RP SEQUENCE FROM N.A.
RC TISSUE=Whole;
RA Klein S., Strausberg R.;
RL Submitted (JUN-2003) to the EMBL/GenBank/DBJ databases.
RW EMBL; BC054306; AAH54306.1;
KW Hypothetical protein.
SQ SEQUENCE 215 AA; 24274 MW; 16B308E7E754F456 CRC64;

Query Match 100.0%; Score 29; DB 13; Length 215;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 SSGPSL 6
Db      157 SSGPSL 162

RESULT 8
Q9FJX4 PRELIMINARY; PRT; 226 AA.
AC Q9FJX4;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Genomic DNA, chromosome 5, TAC clone:K919.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Columbia;
RX MEDLINE=98403884; PubMed=9734815;
RA Kotani H., Nakamura Y., Sato S., Asamizu E., Kaneko T., Miyajima N.,
RA Tabata S.;
RT "Structural analysis of Arabidopsis thaliana chromosome 5. VI.
RL Sequence features of the regions of 1,367,185 bp covered by 19
RT physically assigned P1 and TAC clones.";

```

RL DNA Res. 5:203-216(1998).
 DR EMBL; AB013390; BAB08457.1; -. D833300F9185B978 CRC64;
 SQ SEQUENCE 226 AA; 25019 MW; 25019 MW;
 Query Match 100.0%; Score 29; DB 10; Length 226;
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 SSGPSL 6
 Db 196 SSGPSL 201
 RESULT 9
 QSVKES PRELIMINARY; PRT; 257 AA.
 AC QSVKES;
 DT 01-MAR-2002 (TREMELrel. 20, Created)
 DT 01-MAR-2002 (TREMELrel. 20, Last sequence update)
 DT 01-OCT-2002 (TREMELrel. 22, Last annotation update)
 DE Hypothetical protein MT0781.
 GN MT0781.
 OS Mycobacterium tuberculosis.
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 OX NCBI_TaxID=1773;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CDC 1551 / Oshkosh;
 RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
 RA Peterson J.F., DeBoy R., Dodson R., Gwin M.L., Haft D., Hickey E.,
 RA Kollonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
 RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
 RA Bishai W.;
 RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
 RT laboratory strains."
 RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; A8006969; AAK45022.1; -
 DR TIGR; MT0781; -
 KW Hypothetical protein.
 SQ SEQUENCE 257 AA; 25997 MW; 36D9A9AC973C5D54 CRC64;
 Query Match 100.0%; Score 29; DB 16; Length 257;
 Best Local Similarity 100.0%; Pred. No. 1.5e+02;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 SSGPSL 6
 Db 79 SSGPSL 84
 RESULT 10
 Q8IXVO PRELIMINARY; PRT; 277 AA.
 AC Q8IXVO;
 DT 01-MAR-2003 (TREMELrel. 23, Created)
 DT 01-MAR-2003 (TREMELrel. 23, Last sequence update)
 DT 01-OCT-2003 (TREMELrel. 25, Last annotation update)
 DE Similar to hairy homolog (Drosophila).
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC Straussberg R.;
 RA Straussberg R.;
 RL Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; BC039152; AAH39152.1; -
 DR GO; GO:0003677; F:DNA binding; IEA.
 DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
 DR InterPro; IPR001082; HLH basic.
 DR InterPro; IPR003650; Orange.

DR Pfam; PF00010; HLH; 1.
 DR SMART; SMO0353; HLH; 1.
 DR SMART; SMO0511; ORANGE; 1.
 DR PROSITE; PS00038; HLH; 1; 1.
 DR PROSITE; PS50888; HLH; 2; 1.
 SQ SEQUENCE 277 AA; 29271 MW; FD1B22170CB38338 CRC64;
 Query Match 100.0%; Score 29; DB 4; Length 277;
 Best Local Similarity 100.0%; Pred. No. 1.6e+02;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 SSGPSL 6
 Db 261 SSGPSL 266
 RESULT 11
 Q9L913 PRELIMINARY; PRT; 284 AA.
 AC Q9L913;
 DT 01-OCT-2000 (TREMELrel. 15, Created)
 DT 01-OCT-2000 (TREMELrel. 15, Last sequence update)
 DT 01-JUN-2003 (TREMELrel. 24, Last annotation update)
 DE CopB.
 GN CopB.
 OS Aeromonas veronii bv. sobria.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Aeromonadales;
 OC Aeromonadaceae; Aeromonas.
 OX NCBI_TaxID=114517;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=3787;
 RA Francki K.T., Chang B.J., Mee B.J., Collignon P.J., Keese P.K.;
 RT "An adhesion defective mutant of Aeromonas veronii biovar sobria
 RT associated with copper tolerance."
 RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF187849; AAF32270.1; -
 DR InterPro; IPR007939; CopB.
 DR Pfam; PF05275; CopB; 1.
 SQ SEQUENCE 284 AA; 32401 MW; 68E414EBE3D8F148 CRC64;
 Query Match 100.0%; Score 29; DB 2; Length 284;
 Best Local Similarity 100.0%; Pred. No. 1.7e+02;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 SSGPSL 6
 Db 60 SSGPSL 65
 RESULT 12
 Q8YJL9 PRELIMINARY; PRT; 348 AA.
 AC Q8YJL9;
 DT 01-MAR-2002 (TREMELrel. 20, Created)
 DT 01-MAR-2002 (TREMELrel. 20, Last sequence update)
 DT 01-OCT-2003 (TREMELrel. 25, Last annotation update)
 DE Hypothetical membrane spanning protein BME10063.
 GN BME10063.
 OS Brucella melitensis.
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
 OC Brucellaceae; Brucella.
 OX NCBI_TaxID=29459;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=16M / ATCC 23456 / Biotype 1;
 RX MEDLINE=20020109; PubMed=11756688;
 RA DelVecchio V.G., Kapatral V., Redkar R.J., Patra G., Mijer C., Los T.,
 RA Ivanova N., Anderson I., Bhattacharya A., Lykidis A., Reznik G.,
 RA Jablonski L., Larsen N., D'Souza M., Bernal A., Mazur M., Goltsman E.,
 RA Selkov E., Elzer P.H., Hagius S., O'Callaghan D., Letesson J.-J.,
 RA Haselkorn R., Kyripides N., Overbeek R.;
 RT "The genome sequence of the facultative intracellular pathogen

```

RT Brucella melitensis.;
RL Proc. Natl. Acad. Sci. U.S.A. 99:443-448(2002).
DR EMBL; AE009450; AAL51245.1; -.
DR PIR; AB3260; AB3260.
DR InterPro; IPR008964; Invasin.intim.in.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 348 AA; 36254 MW; 288E042DA7B1774A CRC64;

Query Match 100.0%; Score 29; DB 16; Length 348;
Best Local Similarity 100.0%; Pred. No. 2e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SSGPSL 6
Db 140 SSGPSL 145

RESULT 13
Q7U3U5 PRELIMINARY; PRT; 358 AA.
ID Q7U3U5
AC Q7U3U5
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Aminotransferases Class-I (EC 2.6.1.9).
DE HISC/COBC OR SYNW2334.
GN Synchococcus sp. (strain WH102).
OS Synchococcus.
OC Bacteria; Cyanobacteria; Chroococcales; Synchococcus.
CX NCBI_TaxID=84588;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=22825697; PubMed=12917641;
RA Palenik B., Brahmsha B., Larimer F.W., Land M., Hauser L., Chain P.,
RA Lamerdin J., Regala W., Allen E.E., McCarran J., Paulsen I.,
RA Dufresne A., Partensky F., Webb E.A., Waterbury J.;
RT "The genome of a motile marine Synchococcus."
RL Nature 424:1037-1042(2003).
DR EMBL; BX569695; CAE08849.1; -.
KW Aminotransferase; Transferase; Complete proteome.
SQ SEQUENCE 358 AA; 40325 MW; 0C353C403E54DC7F CRC64;

Query Match 100.0%; Score 29; DB 16; Length 358;
Best Local Similarity 100.0%; Pred. No. 2.1e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SSGPSL 6
Db 92 SSGPSL 97

RESULT 14
Q7XK21 PRELIMINARY; PRT; 392 AA.
ID Q7XK21
AC Q7XK21
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE OSUNBA0044K18.24 protein.
DE OSUNBA0044K18.24.
GN Oryza sativa (Rice).
OS Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzaceae; Oryza.
CX NCBI_TaxID=4530;
RN [1]
RP SEQUENCE FROM N.A.
RX Fu G., Wang S.X., Ren S.X., Lv G., Lin W., Gu W.Q., Zhu G.F., Tu Y.F.,
RA Jia J., Yin H.F., Zhang Y., Cai Z., Chen J., Kang H., Chen X.Y.,
RA Shao Y., Sun Y., Hu Q.P., Zhang X.L., Zhang W., Wang L.J., Ding C.W.,
RA Sheng H.H., Gu J.L., Chen S.T., Ni L., Zhu F.H., Han B., Peng Q.,
RA Huang Y.C., Li Y., Zhu J.J., Zhao Q., Hu X., Lin Y.L., Mu J., Yu Z.,
RA Chen L., Fan D.L., Weng Q.J., Zhang L., Lu Y.Q., Yu S.L., Liu X.H.,
RA Lu T.T., Zhang Y.J., Lu Y., Li C., Li T., Zhang Y., Hu H., Jia P.X.,

```

```

RA Qian Y.M., Ying K., Zhou B., Chen Z.H., Hao P., Zhang L., Wu M.,
RA Zhang R.Q., Guan J.P., Hong G.F.;
RL Submitted (SEP-2001) to the EMBL/GenBank/DBSJ databases.
DR EMBL; AL606595; CAE05882.1; -.
SQ SEQUENCE 392 AA; 42199 MW; 4C7982796D01576F CRC64;

Query Match 100.0%; Score 29; DB 10; Length 392;
Best Local Similarity 100.0%; Pred. No. 2.3e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SSGPSL 6
Db 60 SSGPSL 65

RESULT 15
Q8N4F2 PRELIMINARY; PRT; 395 AA.
ID Q8N4F2
AC Q8N4F2
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE LOC165100.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
CX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX Strausberg R.;
RA Strausberg R.;
RL Submitted (JUL-2002) to the EMBL/GenBank/DBSJ databases.
DR EMBL; BC034405; AAH34405.1; -.
SQ SEQUENCE 395 AA; 41599 MW; E9E4D81DEF45D414 CRC64;

Query Match 100.0%; Score 29; DB 4; Length 395;
Best Local Similarity 100.0%; Pred. No. 2.3e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SSGPSL 6
Db 323 SSGPSL 328

Search completed: March 10, 2004, 09:25:29
Job time : 8.92996 secs

```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 10, 2004, 08:58:53 ; Search time 1.09728 Seconds
(without alignments)
284.724 Million cell updates/sec

Title: US-09-848-834A-6
Perfect score: 29
Sequence: 1 SSGPSL 6

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_42:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	29	100.0	225	1 VV SV41	P36315 simian viru
2	29	100.0	280	1 H51 HUMAN	Q14469 homo sapien
3	29	100.0	395	1 RPP SV41	Q86606 simian viru
4	29	100.0	410	1 HX3 HETFR	Q91a21 heterodontu
5	29	100.0	416	1 NAP5 HUMAN	O14513 homo sapien
6	29	100.0	989	1 FRQ NEUCR	P19970 neurospora
7	29	100.0	997	1 FQO SORFI	Q09033 sordaria fi
8	29	100.0	1149	1 DPOL ADECC	Q65946 canine aden
9	29	100.0	1150	1 DPOL ADECT	P87553 canine aden
10	29	100.0	1210	1 BAT8 HUMAN	Q96kq7 homo sapien
11	29	100.0	1263	1 BAT8 MOUSE	Q92148 mus musculu
12	29	100.0	1380	1 ZMS1 YEAST	P46974 saccharomyc
13	29	100.0	5560	1 SPEN DROME	Q8sx83 drosophila
14	27	93.1	526	1 KCH2 CHICK	Q9pt84 gallus gall
15	27	93.1	935	1 MYCD MOUSE	Q8vims mus musculu
16	27	93.1	938	1 MYCD RAT	Q8v517 rattus norv
17	27	93.1	1195	1 KCH7 MOUSE	Q9er47 mus musculu
18	27	93.1	1195	1 KCH7 RAT	O54852 rattus norv
19	27	93.1	1196	1 KCH7 HUMAN	Q9ns40 homo sapien
20	27	93.1	2174	1 T240 HUMAN	Q9uhv7 homo sapien
21	26	89.7	102	1 CYC EUGGR	P00076 euglena gra
22	26	89.7	209	1 HLA XENLA	P04892 xenopus lae
23	26	89.7	297	1 YU8O CAEL	Q93591 caenorhabdi
24	26	89.7	299	1 CRX BOVIN	Q9xsx0 bos taurus
25	26	89.7	299	1 CRX HUMAN	O43186 homo sapien
26	26	89.7	299	1 CRX MOUSE	O54751 mus musculu
27	26	89.7	340	1 PURA METTH	O26712 methanobact
28	26	89.7	366	1 LEU3 BACCA	P05644 bacillus ca
29	26	89.7	392	1 PSME SULTO	Q97su2 sulfolobus
30	26	89.7	393	1 PSME SULTO	Q980m1 sulfolobus
31	26	89.7	399	1 HXB3 CHICK	P23682 gallus gall
32	26	89.7	457	1 NIFE KLEPN	P08737 klebsiella
33	26	89.7	496	1 NIFE RHIL	Q98ap4 rhizobium 1

US-09-848-834A-6

141681 seqs, 52070155 residues

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_42:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	29	100.0	225	1 VV SV41	P36315 simian viru
2	29	100.0	280	1 H51 HUMAN	Q14469 homo sapien
3	29	100.0	395	1 RPP SV41	Q86606 simian viru
4	29	100.0	410	1 HX3 HETFR	Q91a21 heterodontu
5	29	100.0	416	1 NAP5 HUMAN	O14513 homo sapien
6	29	100.0	989	1 FRQ NEUCR	P19970 neurospora
7	29	100.0	997	1 FQO SORFI	Q09033 sordaria fi
8	29	100.0	1149	1 DPOL ADECC	Q65946 canine aden
9	29	100.0	1150	1 DPOL ADECT	P87553 canine aden
10	29	100.0	1210	1 BAT8 HUMAN	Q96kq7 homo sapien
11	29	100.0	1263	1 BAT8 MOUSE	Q92148 mus musculu
12	29	100.0	1380	1 ZMS1 YEAST	P46974 saccharomyc
13	29	100.0	5560	1 SPEN DROME	Q8sx83 drosophila
14	27	93.1	526	1 KCH2 CHICK	Q9pt84 gallus gall
15	27	93.1	935	1 MYCD MOUSE	Q8vims mus musculu
16	27	93.1	938	1 MYCD RAT	Q8v517 rattus norv
17	27	93.1	1195	1 KCH7 MOUSE	Q9er47 mus musculu
18	27	93.1	1195	1 KCH7 RAT	O54852 rattus norv
19	27	93.1	1196	1 KCH7 HUMAN	Q9ns40 homo sapien
20	27	93.1	2174	1 T240 HUMAN	Q9uhv7 homo sapien
21	26	89.7	102	1 CYC EUGGR	P00076 euglena gra
22	26	89.7	209	1 HLA XENLA	P04892 xenopus lae
23	26	89.7	297	1 YU8O CAEL	Q93591 caenorhabdi
24	26	89.7	299	1 CRX BOVIN	Q9xsx0 bos taurus
25	26	89.7	299	1 CRX HUMAN	O43186 homo sapien
26	26	89.7	299	1 CRX MOUSE	O54751 mus musculu
27	26	89.7	340	1 PURA METTH	O26712 methanobact
28	26	89.7	366	1 LEU3 BACCA	P05644 bacillus ca
29	26	89.7	392	1 PSME SULTO	Q97su2 sulfolobus
30	26	89.7	393	1 PSME SULTO	Q980m1 sulfolobus
31	26	89.7	399	1 HXB3 CHICK	P23682 gallus gall
32	26	89.7	457	1 NIFE KLEPN	P08737 klebsiella
33	26	89.7	496	1 NIFE RHIL	Q98ap4 rhizobium 1

US-09-848-834A-6

141681 seqs, 52070155 residues

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_42:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	29	100.0	225	1 VV SV41	P36315 simian viru
2	29	100.0	280	1 H51 HUMAN	Q14469 homo sapien
3	29	100.0	395	1 RPP SV41	Q86606 simian viru
4	29	100.0	410	1 HX3 HETFR	Q91a21 heterodontu
5	29	100.0	416	1 NAP5 HUMAN	O14513 homo sapien
6	29	100.0	989	1 FRQ NEUCR	P19970 neurospora
7	29	100.0	997	1 FQO SORFI	Q09033 sordaria fi
8	29	100.0	1149	1 DPOL ADECC	Q65946 canine aden
9	29	100.0	1150	1 DPOL ADECT	P87553 canine aden
10	29	100.0	1210	1 BAT8 HUMAN	Q96kq7 homo sapien
11	29	100.0	1263	1 BAT8 MOUSE	Q92148 mus musculu
12	29	100.0	1380	1 ZMS1 YEAST	P46974 saccharomyc
13	29	100.0	5560	1 SPEN DROME	Q8sx83 drosophila
14	27	93.1	526	1 KCH2 CHICK	Q9pt84 gallus gall
15	27	93.1	935	1 MYCD MOUSE	Q8vims mus musculu
16	27	93.1	938	1 MYCD RAT	Q8v517 rattus norv
17	27	93.1	1195	1 KCH7 MOUSE	Q9er47 mus musculu
18	27	93.1	1195	1 KCH7 RAT	O54852 rattus norv
19	27	93.1	1196	1 KCH7 HUMAN	Q9ns40 homo sapien
20	27	93.1	2174	1 T240 HUMAN	Q9uhv7 homo sapien
21	26	89.7	102	1 CYC EUGGR	P00076 euglena gra
22	26	89.7	209	1 HLA XENLA	P04892 xenopus lae
23	26	89.7	297	1 YU8O CAEL	Q93591 caenorhabdi
24	26	89.7	299	1 CRX BOVIN	Q9xsx0 bos taurus
25	26	89.7	299	1 CRX HUMAN	O43186 homo sapien
26	26	89.7	299	1 CRX MOUSE	O54751 mus musculu
27	26	89.7	340	1 PURA METTH	O26712 methanobact
28	26	89.7	366	1 LEU3 BACCA	P05644 bacillus ca
29	26	89.7	392	1 PSME SULTO	Q97su2 sulfolobus
30	26	89.7	393	1 PSME SULTO	Q980m1 sulfolobus
31	26	89.7	399	1 HXB3 CHICK	P23682 gallus gall
32	26	89.7	457	1 NIFE KLEPN	P08737 klebsiella
33	26	89.7	496	1 NIFE RHIL	Q98ap4 rhizobium 1

US-09-848-834A-6

141681 seqs, 52070155 residues

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_42:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	29	100.0	225	1 VV SV41	P36315 simian viru
2	29	100.0	280	1 H51 HUMAN	Q14469 homo sapien
3	29	100.0	395	1 RPP SV41	Q86606 simian viru
4	29	100.0	410	1 HX3 HETFR	Q91a21 heterodontu
5	29	100.0	416	1 NAP5 HUMAN	O14513 homo sapien
6	29	100.0	989	1 FRQ NEUCR	P19970 neurospora
7	29	100.0	997	1 FQO SORFI	Q09033 sordaria fi
8	29	100.0	1149	1 DPOL ADECC	Q65946 canine aden
9	29	100.0	1150	1 DPOL ADECT	P87553 canine aden
10	29	100.0	1210	1 BAT8 HUMAN	Q96kq7 homo sapien
11	29	100.0	1263	1 BAT8 MOUSE	Q92148 mus musculu
12	29	100.0	1380	1 ZMS1 YEAST	P46974 saccharomyc
13	29	100.0	5560	1 SPEN DROME	Q8sx83 drosophila
14	27	93.1	526	1 KCH2 CHICK	Q9pt84 gallus gall
15	27	93.1	935	1 MYCD MOUSE	Q8vims mus musculu
16	27	93.1	938	1 MYCD RAT	Q8v517 rattus norv
17	27	93.1	1195	1 KCH7 MOUSE	Q9er47 mus musculu
18	27	93.1	1195	1 KCH7 RAT	O54852 rattus norv
19	27	93.1	1196	1 KCH7 HUMAN	Q9ns40 homo sapien
20	27	93.1	2174	1 T240 HUMAN	Q9uhv7 homo sapien
21	26	89.7	102	1 CYC EUGGR	P00076 euglena gra
22	26	89.7	209	1 HLA XENLA	P04892 xenopus lae
23	26	89.7	297	1 YU8O CAEL	Q93591 caenorhabdi
24	26	89.7	299	1 CRX BOVIN	Q9xsx0 bos taurus
25	26	89.7	299	1 CRX HUMAN	O43186 homo sapien
26	26	89.7	299	1 CRX MOUSE	O54751 mus musculu
27	26	89.7	340	1 PURA METTH	O26712 methanobact
28	26	89.7	366	1 LEU3 BACCA	P05644 bacillus ca
29	26	89.7	392	1 PSME SULTO	Q97su2 sulfolobus
30	26	89.7	393	1 PSME SULTO	Q980m1 sulfolobus
31	26	89.7	399	1 HXB3 CHICK	P23682 gallus gall
32	26	89.7	457	1 NIFE KLEPN	P08737 klebsiella
33	26	89.7	496	1 NIFE RHIL	Q98ap4 rhizobium 1

HES1_HUMAN STANDARD; PRT; 280 AA.

AC Q14469; 15-DEC-1998 (Rel. 37, Created)

DT 15-DEC-1998 (Rel. 37, Last sequence update)

DT 15-MAR-2004 (Rel. 43, Last annotation update)

DE Transcription factor HES-1 (Hairy and enhancer of split 1) (Hairy-like) (HLH) (Hairy homolog).

DE HES1 OR HRY OR HLH

GN Homo sapiens (Human)

OS Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=94232187; PubMed=8020957;

RA Feder J.N., Li L., Jan L.Y., Jan Y.-N.;

RT "Genomic cloning and chromosomal localization of HRY, the human homolog to the Drosophila segmentation gene, hairy.";

RL Genomics 20:56-61(1994).

RN [2]

RP SEQUENCE FROM N.A.

RC TISSUE=Kidney;

RA Yao J., Yeung S., Sun H., Chen N.;

RT "Functional analysis of human HRY in Drosophila.";

RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.

RN [3]

RP SEQUENCE FROM N.A.

RA Watanabe K., Kumagai A., Itakura S., Yamazaki M., Tashiro H., Ota T., Suzuki Y., Oabayashi M., Nishi T., Shibahara T., Tanaka T., Nakamura Y., Iacgai T., Sugano S.;

RT "NEDO human cDNA sequencing project.";

RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.

RN [4]

RP INTERACTION WITH SIRT1.

RX MEDLINE=22422590; PubMed=12535671;

RA Takata T., Ishikawa F.;

RT "Human Sir2-related protein SIRT1 associates with the bHLH repressors HES1 and HEV2 and is involved in HES1- and HEV2-mediated transcriptional repression.";

RL Biochem. Biophys. Res. Commun. 301:250-257(2003).

CC -!- FUNCTION: Transcriptional repressor of genes that require a bHLH protein for their transcription. May act as a negative regulator of myogenesis by inhibiting the functions of MYO1 and ASH1 (By similarity).

CC -!- SUBUNIT: Transcription repression requires formation of a complex with a co-repressor protein (Groucho) such as TLE2 (By similarity). Interacts with SIRT1.

CC -!- SUBCELLULAR LOCATION: Nuclear.

CC -!- DOMAIN: Has a particular type of basic domain (presence of a helix-interrupting proline) that binds to the N-box (CACNAG), rather than the canonical E-box (CANNTG).

CC -!- DOMAIN: The carboxyl-terminal WRPW motif is a transcriptional repression domain necessary for the interaction with Groucho, a transcriptional co-repressor recruited to specific target DNA by Hairy-related proteins.

CC -!- SIMILARITY: Contains 1 basic helix-loop-helix (bHLH) domain.

CC -!- SIMILARITY: Contains 1 orange domain.

CC

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC

CC EMBL; L19314; AAA65220.1; -

CC EMBL; AF24785; AAF73060.1; -

CC EMBL; AK000415; BAA91149.1; -

CC PIR; A53027; A53027.

CC DR TRANSFAC; T04892; -

CC Genew; HGNC:5192; HES1.

DR

DR MIM; 139605; -

DR GO; GO:0003677; F:DNA binding; TAS.

DR GO; GO:0007399; P:neurogenesis; TAS.

DR InterPro; IPR001092; HLH_basic.

DR InterPro; IPR003650; Orange.

DR Pfam; PF00010; HLH; 1.

DR SMART; SM00353; HLH; 1.

DR SMART; SM00511; ORANGE; 1.

DR PROSITE; PSS0888; HLH; 1.

KW Transcription regulation; DNA-binding; Nuclear protein; Repressor.

FT DNA_BIND 35 47

FT DOMAIN 48 92

FT DOMAIN 156 245

FT DOMAIN 249 273

FT DOMAIN 275 278

FT

SQ SEQUENCE 280 AA; 29541 MW; F9342A88FC749E3C CRC64;

Query Match 100.0%; Score 29; DB 1; Length 280;

Best Local Similarity 100.0%; Pred. No. 37;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Cy 1 SSGPSL 6

Db 264 SSGPSL 269

RESULT 3

RRPP SV41 STANDARD; PRT; 395 AA.

ID RRPP SV41

AC Q86606;

DT 15-MAR-2004 (Rel. 43, Created)

DT 15-MAR-2004 (Rel. 43, Last sequence update)

DT 15-MAR-2004 (Rel. 43, Last annotation update)

DE RNA polymerase alpha subunit (EC 2.7.7.48) (p protein).

GN P.V.

OS Simian virus 41 (SV41).

OC Viruses; ssRNA negative-strand viruses; Mononegavirales; Paramyxoviridae; Paramyxovirinae; Rubulavirus.

OX NCBI_TaxID=11228;

RN [1]

RP SEQUENCE FROM N.A. AND RNA EDITING.

RC STRAIN=Toshiha/Chanock;

RX MEDLINE=93260408; PubMed=8492098;

RA Kawano M., Teuridome M., Oki N., Nishio M., Komada H., Matsumura H., Kusagawa S., Ohta H., Ito Y.;

RT "Sequence determination of the P gene of simian virus 41: presence of irregular deletions near the RNA-editing sites of paramyxoviruses.";

RL J. Gen. Virol. 74:911-916(1993).

CC -!- FUNCTION: Essential component of the RNA polymerase and the nascent chain assembly complex. Also required during RNA synthesis.

CC -!- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate + (RNA)(N).

CC -!- RNA EDITING: Modified positions=164; Notes=Partially edited. RNA editing at this position consists of an insertion of two guanine nucleotides. The sequence displayed here is the P protein, derived from the edited RNA. The unedited RNA version gives rise to the V protein (AC P36315).

CC

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC

CC EMBL; S60813; AAB26640.2; -

CC PIR; JQ2041; JQ2041.

CC InterPro; IPR004897; Paramyx_P_V.

CC Pfam; PF03210; Paramyx_P_V; 1.

CC Transferase; RNA-directed RNA polymerase; Nucleocapsid;

KW

KW Phosphorylation; RNA editing.

SQ SEQUENCE 395 AA; 41920 MW; 6B97477BF483D30 CRC64;

Query Match 100.0%; Score 29; DB 1; Length 395;

Best Local Similarity 100.0%; Pred. No. 52;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SSGPSL 6

Db 144 SSGPSL 149

RESULT 4

HXA3_HETPR

ID HXA3_HETPR STANDARD; PRT; 410 AA.

AC Q91A21;

DT 16-OCT-2001 (Rel. 40, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE Homeobox protein Hox-A3.

GN HXA3.

OS Heterodontus francisci (Horn shark).

OC Eukaryota; Metazoa; Chordata; Vertebrata; Chondrichthyes;

OC Elasmobranchii; Galeomorphii; Heterodontidae; Heterodontiformes;

OC Heterodontidae; Heterodontus.

OX NCBI_TaxID=7792;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=20144096; PubMed=10677514;

RA Kim C.B., Amemiya C., Bailey W., Kawasaki K., Mezey J., Miller W.,

RA Minoshima S., Shimizu N., Wagner G., Ruddle F.,

RT "Hox cluster genomics in the horn shark, Heterodontus francisci.";

RL Proc. Natl. Acad. Sci. U.S.A. 97:1655-1660(2000).

CC -!- FUNCTION: Sequence-specific transcription factor which is part of

CC a developmental regulatory system that provides cells with

CC specific positional identities on the anterior-posterior axis (By

CC similarity).

CC -!- SUBCELLULAR LOCATION: Nuclear (By similarity).

CC -!- SIMILARITY: Belongs to the Antp homeobox family.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>CC or send an email to license@isb-sib.ch).

CC -----

DR EMBL; AF224262; AAF44641.1; --

DR HSSP; P02833; ISAN.

DR TRANSFAC; T04474; --

DR InterPro; IPR001827; Antennapedia.

DR InterPro; IPR001356; Homeobox.

DR Pfam; PF00046; homeobox; 1.

DR PRINTS; PR00025; ANTENNAPEDIA.

DR PRINTS; PR00024; HOMEBOX.

DR ProDom; PD000010; Homeobox; 1.

DR SMART; SM00389; HOX; 1.

DR PROSITE; PS00032; ANTENNAPEDIA; 1.

DR PROSITE; PS00027; HOMEBOX 1; 1.

DR PROSITE; PS00071; HOMEBOX 2; 1.

KW Homeobox; DNA-binding; Developmental protein; Nuclear protein;

KW Transcription regulation.

KW SITE 133 138 ANTP-TYPE HEXAPEPTIDE.

FT DNA_BIND 168 227 HOMEBOX.

FT DOMAIN 85 90 POLY-GLN.

FT DOMAIN 91 100 POLY-PRO.

FT DOMAIN 149 153 POLY-SER.

FT SEQUENCE 410 AA; 44548 MW; 285ABC06B41C9FD9 CRC64;

Query Match

Best Local Similarity 100.0%; Score 29; DB 1; Length 410;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SSGPSL 6

Db 346 SSGPSL 351

RESULT 5

NAP5_HUMAN

ID NAP5_HUMAN STANDARD; PRT; 416 AA.

AC O14513;

DT 28-FEB-2003 (Rel. 41, Created)

DT 28-FEB-2003 (Rel. 41, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE Nck-associated protein 5 (NAP-5) (Fragment).

GN NAP5.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.

OX NCBI_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=Brain;

RX MEDLINE=98008666; PubMed=9344857;

RA Matuoka K., Miki H., Takahashi K., Takenawa T.;

RT "A novel ligand for an SH3 domain of the adaptor protein Nck bears an

RT SH2 domain and nuclear signaling motifs.";

RL Biochem. Biophys. Res. Commun. 239:488-492(1997).

CC -!- SUBUNIT: Interacts with the SH3-containing region of the adaptor

CC protein Nck

CC -!- TISSUE SPECIFICITY: Expressed in fetal and adult brain, leukocytes

CC and fetal fibroblasts.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>CC or send an email to license@isb-sib.ch).

CC -----

DR EMBL; AB005217; BAA22433.1; --

FT NON_TER 1 1

FT NON_TER 416 416

SQ SEQUENCE 416 AA; 44829 MW; AB4B7F3A19522612 CRC64;

Query Match

Best Local Similarity 100.0%; Score 29; DB 1; Length 416;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SSGPSL 6

Db 192 SSGPSL 197

RESULT 6

FRQ_NEUCR

ID FRQ_NEUCR STANDARD; PRT; 989 AA.

AC P19970; Q01276.

DT 01-FEB-1991 (Rel. 17, Created)

DT 15-JUL-1999 (Rel. 38, Last sequence update)

DT 15-MAR-2004 (Rel. 43, Last annotation update)

DE Frequency clock protein.

GN FRQ.

OS Neurospora crassa.

OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;

OC Sordariomycetidae; Sordariales; Sordariaceae; Neurospora.

OX NCBI_TaxID=5141;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=74-OR23-1A / FGSC 987;

RX MEDLINE=94329580; PubMed=8052643;

RA Aronson B.D. Johnson K.A., Dunlap J.C.;

RT "Circadian clock locus frequency: protein encoded by a single open

RT reading frame defines period length and temperature compensation.";
RL Proc. Natl. Acad. Sci. U.S.A. 91:7683-7687(1994).
RN [2]
RP SEQUENCE OF 202-989 FROM N.A.
RX MEDLINE=89281721; PubMed=2525233;
RA McClung C.R., Fox B.A., Dunlap J.C.;
RT "The Neurospora clock gene frequency shares a sequence element with
the *Drosophila* clock gene period.";
RL Nature 339:558-562(1989).
RN [3]
RP PHOSPHORYLATION BY CK2.
RX MEDLINE=21956574; PubMed=11959847;
RA Yang Y., Cheng P., Liu Y.;
RT "Regulation of the *Neurospora* circadian clock by casein kinase II.";
RL Genes Dev. 16:994-1006(2002).
RN [4]
RP FUNCTION: Circadian clock component involved in the generation of
biological rhythms, in particular in rhythm stability, period
length, and temperature compensation. Oscillates in abundance with
a daily peak early in the morning. Behaves as a negative element
in circadian transcriptional loop. May bind to wc2 protein. The
complex frq-wc2 may turn off the expression of frq.
CC -1- ALTERNATIVE PRODUCTS:
CC Event-Alternative initiation:
CC Comment=2 isoforms Long (shown here) and Short, are produced by
alternative initiation. The isoform Long maintains rhythms at
high temperature (30 degrees Celsius), while the isoform Short
maintains rhythms at lower temperature (18 degrees Celsius);
CC -1- INDUCTION: By light; perhaps through white collar-1 (wcl) and
white collar-2 (wc2). Also activated directly by wcl and wc2.
CC -1- PTM: Progressive phosphorylation is also involved in regulating frq
early night. Phosphorylation by CKII may have at least three
degradation. Phosphorylation by CKII reduces the protein
functions; it decreases the stability of frq, reduces the protein
complex formation between frq and the white collar proteins, and
is important for the closing of the *Neurospora* circadian negative
feedback loop.
CC -1- SIMILARITY: BELONGS TO THE FRQ FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U17073; AAA57121.1; --
DR PIR; T46659; T46659.
KW Biological rhythms; Transcription regulation; Nuclear protein;
KW Phosphorylation; Alternative initiation.
FT CHAIN 1 989 FREQUENCY CLOCK PROTEIN, ISOFORM LONG.
FT CHAIN 1 989 FREQUENCY CLOCK PROTEIN, ISOFORM SHORT.
FT INIT MET 100 100 FOR ISOFORM SHORT.
FT DOMAIN 244 250 POLY-SER.
FT DOMAIN 564 568 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT DOMAIN 862 888 ASP/GLU-RICH (ACIDIC).
FT MUTAGEN 364 369 E->K: IN FRQ3; LENGTHENED PERIOD.
FT MUTAGEN 459 459 G->D: IN FRQ7; LENGTHENED PERIOD.
FT MUTAGEN 482 482 G->S: IN FRQ1; SHORTENED PERIOD.
FT MUTAGEN 663 989 MISSING: IN FRQ9; LOSS OF RHYTHMICITY.
FT MUTAGEN 885 895 A->T: IN FRQ2; SHORTENED PERIOD.
FT MUTAGEN 202 207 EATLH -> NGGRP (IN REF. 2).
FT CONFLICT 233 233 Y -> H (IN REF. 2).
SQ SEQUENCE 989 AA; 109229 MW; 79E2A353E6A8099F CRC64;
Query Match 100.0%; Score 29; DB 1; Length 989;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 SSGPSL 6
DB 271 SSGPSL 276

RT reading frame defines period length and temperature compensation.";
RL Proc. Natl. Acad. Sci. U.S.A. 91:7683-7687(1994).
RN [2]
RP SEQUENCE OF 202-989 FROM N.A.
RX MEDLINE=89281721; PubMed=2525233;
RA McClung C.R., Fox B.A., Dunlap J.C.;
RT "The Neurospora clock gene frequency shares a sequence element with
the *Drosophila* clock gene period.";
RL Nature 339:558-562(1989).
RN [3]
RP PHOSPHORYLATION BY CK2.
RX MEDLINE=21956574; PubMed=11959847;
RA Yang Y., Cheng P., Liu Y.;
RT "Regulation of the *Neurospora* circadian clock by casein kinase II.";
RL Genes Dev. 16:994-1006(2002).
RN [4]
RP FUNCTION: Circadian clock component involved in the generation of
biological rhythms, in particular in rhythm stability, period
length, and temperature compensation. Oscillates in abundance with
a daily peak early in the morning. Behaves as a negative element
in circadian transcriptional loop. May bind to wc2 protein. The
complex frq-wc2 may turn off the expression of frq.
CC -1- ALTERNATIVE PRODUCTS:
CC Event-Alternative initiation:
CC Comment=2 isoforms Long (shown here) and Short, are produced by
alternative initiation. The isoform Long maintains rhythms at
high temperature (30 degrees Celsius), while the isoform Short
maintains rhythms at lower temperature (18 degrees Celsius);
CC -1- INDUCTION: By light; perhaps through white collar-1 (wcl) and
white collar-2 (wc2). Also activated directly by wcl and wc2.
CC -1- PTM: Progressive phosphorylation is also involved in regulating frq
early night. Phosphorylation by CKII may have at least three
degradation. Phosphorylation by CKII reduces the protein
functions; it decreases the stability of frq, reduces the protein
complex formation between frq and the white collar proteins, and
is important for the closing of the *Neurospora* circadian negative
feedback loop.
CC -1- SIMILARITY: BELONGS TO THE FRQ FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U17073; AAA57121.1; --
DR PIR; T46659; T46659.
KW Biological rhythms; Transcription regulation; Nuclear protein;
KW Phosphorylation; Alternative initiation.
FT CHAIN 1 989 FREQUENCY CLOCK PROTEIN, ISOFORM LONG.
FT CHAIN 1 989 FREQUENCY CLOCK PROTEIN, ISOFORM SHORT.
FT INIT MET 100 100 FOR ISOFORM SHORT.
FT DOMAIN 244 250 POLY-SER.
FT DOMAIN 564 568 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT DOMAIN 862 888 ASP/GLU-RICH (ACIDIC).
FT MUTAGEN 364 369 E->K: IN FRQ3; LENGTHENED PERIOD.
FT MUTAGEN 459 459 G->D: IN FRQ7; LENGTHENED PERIOD.
FT MUTAGEN 482 482 G->S: IN FRQ1; SHORTENED PERIOD.
FT MUTAGEN 663 989 MISSING: IN FRQ9; LOSS OF RHYTHMICITY.
FT MUTAGEN 885 895 A->T: IN FRQ2; SHORTENED PERIOD.
FT MUTAGEN 202 207 EATLH -> NGGRP (IN REF. 2).
FT CONFLICT 233 233 Y -> H (IN REF. 2).
SQ SEQUENCE 989 AA; 109229 MW; 79E2A353E6A8099F CRC64;
Query Match 100.0%; Score 29; DB 1; Length 989;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 SSGPSL 6
DB 271 SSGPSL 276

RESULT 7
FRQ_SORFI STANDARD; PRT; 997 AA.
ID FRQ_SORFI
AC Q09033;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Frequency clock protein.
GN FRQ.
OS Sordaria fimicola.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Sordariomycetidae; Sordariales; Sordariaceae; Sordaria.
OX NCBI_TaxID=27338;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=FGSC 2918; PubMed=8194516;
MEDLINE=94252311; PubMed=8194516;
RA Merrow M.W., Dunlap J.C.;
RT "Intergenic complementation of a circadian rhythmicity defect:
phylogenetic conservation of structure and function of the clock gene
frequency.";
RL EMBL J. 13:2257-2266(1994).
CC -1- FUNCTION: Circadian clock component involved in the generation of
biological rhythms, in particular in rhythm stability, period
length, and temperature compensation. Behaves as a negative
element in circadian transcriptional loop.
CC -1- SIMILARITY: BELONGS TO THE FRQ FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
or send an email to license@isb-sib.ch).
CC -----
CC EMBL; L14467; AAA20825.1; --
DR PIR; S44457; S44457.
KW Biological rhythms; Transcription regulation; Nuclear protein.
FT DOMAIN 250 256 POLY-SER.
FT DOMAIN 569 573 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT DOMAIN 868 894 ASP/GLU-RICH (ACIDIC).
SQ SEQUENCE 997 AA; 109056 MW; 2581DD2C62DFF3 CRC64;
Query Match 100.0%; Score 29; DB 1; Length 997;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 SSGPSL 6
DB 277 SSGPSL 282

RESULT 8
DPOL ADECC STANDARD; PRT; 1149 AA.
ID DPOL ADECC
AC Q65946;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE DNA polymerase (EC 2.7.7.7).
GN POL
OS Canine adenovirus type 1 (strain CLL).
OC Viruses; dsDNA viruses, no RNA stage; Adenoviridae; Mastadenovirus.
OX NCBI_TaxID=69150;
RN [1]
RP SEQUENCE FROM N.A.
RA Campbell J.B., Zhao Y.;
RT Submitted (AUG-1996) to the EMBL/GenBank/DBS databases.
CC -1- CATALYTIC ACTIVITY: N deoxynucleoside triphosphate = N diphosphate
+ {DNA}(N)
CC -1- MISCELLANEOUS: This DNA polymerase requires a protein as a primer.
CC -----

CC -|- SIMILARITY: Belongs to the DNA polymerase type-B family.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

DR EMBL; U55001; AAB05434.1; -
 DR InterPro; IPR006172; DNA_pol_B.
 DR InterPro; IPR004868; DNA_pol_B_2.
 DR InterPro; IPR004137; Prismae.
 DR Pfam; PF03175; DNA_pol_B_2; 1.
 DR PRINTS; PR00106; DNAPOLB.
 DR SMART; SM00486; POLBc; 1.
 DR PROSITE; PS00116; DNA_POLYMERASE_B; 1.
 DR TRANSFERASE; DNA-directed DNA polymerase; DNA replication;
 KW DNA-binding.
 SQ SEQUENCE 1149 AA; 130939 MW; A0B686FAEFPA236E CRC64;

Query Match 100.0%; Score 29; DB 1; Length 1149;
 Best Local Similarity 100.0%; Pred. No. 1.6e+02;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SSGPSL 6
 Dd 25 SSGPSL 30

RESULT 9
 ID DPOL ADECT STANDARD; PRT; 1150 AA.
 AC P87553.
 DT 15-DEC-1998 (Rel. 37, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE DNA polymerase (EC 2.7.7.7).
 GN POL.
 OS Canine adenovirus type 2 (strain Toronto A 26-61).
 OC Viruses; dsDNA viruses, no RNA stage; Adenoviridae; Mastadenovirus.
 OX NCBI_TaxID=69152;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Campbell J.B., Zhao Y.;
 RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.
 CC -|- CATALYTIC ACTIVITY: N deoxynucleoside triphosphate = N diphosphate
 CC + [DNA](N).
 CC -|- MISCELLANEOUS: This DNA polymerase requires a protein as a primer.
 CC -|- SIMILARITY: Belongs to the DNA polymerase type-B family.
 CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

DR EMBL; U77082; AAB38716.1; -
 DR InterPro; IPR006172; DNA_pol_B.
 DR InterPro; IPR004868; DNA_pol_B_2.
 DR InterPro; IPR004137; Prismae.
 DR Pfam; PF03175; DNA_pol_B_2; 1.
 DR PRINTS; PR00106; DNAPOLB.
 DR SMART; SM00486; POLBc; 1.
 DR PROSITE; PS00116; DNA_POLYMERASE_B; 1.
 DR TRANSFERASE; DNA-directed DNA polymerase; DNA replication;
 KW DNA-binding.
 SQ SEQUENCE 1150 AA; 130459 MW; 474E0F53705C5873 CRC64;

Query Match 100.0%; Score 29; DB 1; Length 1150;

Best Local Similarity 100.0%; Pred. No. 1.6e+02;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SSGPSL 6
 Dd 25 SSGPSL 30

RESULT 10
 BAT8 HUMAN
 ID BAT8 HUMAN STANDARD; PRT; 1210 AA.
 AC Q96KQ7; Q14349; Q96MH5; Q96QD0; Q9UQL8; Q9Y331;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Histone-lysine N-methyltransferase, H3 lysine-9 specific 3
 DE (EC 2.1.1.43) (Histone H3-K9 methyltransferase 3) (H3-K9-HMTase 3)
 DE (HLA-B associated transcript 8) (G9a) (NG336).
 GN BAT8 OR G9A OR NG336.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Rowen L., Qin S., Madan A., Dickhoff R., Dors M., Madan A., Hicks P.,
 RA Loretz C., Ratcliffe A., Abbasi N., Shaffer T., Hood L.;
 RT "Sequence of the human major histocompatibility complex class III
 RT region".
 RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Hirakawa M., Yamaguchi H., Imai K., Shimada J., Shina S., Tamiya G.,
 RA Oka A., Inoko H.;
 RT "Homo sapiens 2,229,817bp genomic DNA of 6p21.3 HLA class I region."
 RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.

RP SEQUENCE FROM N.A. (ISOFORM 1), ALTERNATIVE SPLICING (ISOFORM 2), AND
 TISSUE SPECIFICITY.
 RX MEDLINE=21564388; PubMed=11707778;
 RA Brown S.E., Campbell R.D., Sanderson C.M.;
 RT "Novel NG36/G9a gene products encoded within the human and mouse MHC
 RT class III regions".
 RL Mamm. Genome 12:916-924 (2001).
 RN [4]
 RP SEQUENCE FROM N.A. (ISOFORM 3).
 RC TISSUE=Salivary Gland;
 RA Oshima A., Takahashi-Fujii A., Tanase T., Imose N., Takeuchi K.,
 RA Arita M., Musashino K., Yuuki H., Hara H., Sugiyama T., Irie R.,
 RA Otsuki T., Sato H., Wakamatsu A., Ishii S., Yamamoto J., Isono Y.,
 RA Kawai-Hio Y., Saito K., Nishikawa T., Kimura K., Yamashita H.,
 RA Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Kanda K., Wagatsuma M.,
 RA Murakawa K., Kanenori K., Sugiyama A., Kawakami B., Suzuki Y.,
 RA Sugano S., Nagahori K., Masuho Y., Nagai K., Isogai T.;
 RT "NEDO human cDNA sequencing project".
 RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
 RN [5]
 RP SEQUENCE OF 5-1210 FROM N.A. (ISOFORM 1).
 RC TISSUE=Muscle, and Uterus;
 RX MEDLINE=22388257; PubMed=12477932;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Scapleton M., Soares M.B., Donald M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Uedin T.S., Tomilki S., Carninci P., Prange C.,
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahey J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
 RA Whitting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,

the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; AF134726; AAD21811.1; ALT SEQ.
EMBL; AF134726; AAD21812.1; ALT SEQ.
EMBL; AF000502; BAB63294.1; ALT SEQ.
EMBL; AF000502; BAB63295.1; ALT SEQ.
EMBL; AJ315532; CAC86666.1; -
EMBL; AK056936; BAB71314.1; -
EMBL; BC002686; AAH02686.1; ALT INIT.
EMBL; BC018718; AAH18718.1; ALT INIT.
EMBL; BC020970; AAH20970.1; ALT INIT.
EMBL; X69838; CAA49491.1; ALT INIT.
MIM; 604599; -
InterPro; IPR002110; ANK.
InterPro; IPR003616; PostSET.
InterPro; IPR007728; Pre-SET.
InterPro; IPR001214; SET.
InterPro; IPR003606; Zr2-binding.
Pfam; PF00023; ank; 6.
Pfam; PF05033; Pre-SET; 1.
Pfam; PF00856; SET; 1.
PRINTS; PR01415; ANKYRIN.
SMART; SMC0248; ANK; 6.
SMART; SMC0468; PreSET; 1.
SMART; SMC0317; SET; 1.
PROSITE; PSS0088; ANK REPEAT; 5.
PROSITE; PSS0297; ANK REP REGION; 1.
PROSITE; PSS0868; POST SET; FALSE-NEG.
PROSITE; PSS0867; PRE SET; 1.
PROSITE; PSS0280; SET; 1.
Transferase; Methyltransferase; Chromatin regulator; Nuclear protein; ANK repeat; Repeat; Alternative splicing.
DOMAIN 2 13 POLY-ALA.
DOMAIN 160 163 POLY-GLU.
DOMAIN 300 326 ANK 1.
REPEAT 649 678 ANK 2.
REPEAT 684 713 ANK 3.
REPEAT 717 746 ANK 4.
REPEAT 750 780 ANK 5.
REPEAT 784 813 ANK 6.
REPEAT 817 846 ANK 7.
REPEAT 850 879 SET.
DOMAIN 972 1035 POST-SET.
DOMAIN 1037 1159 SET.
DOMAIN 1164 1180 Missing (in isoform 2).
VARSPPLIC 373 406 /FTID=VSP_002211.
FT VARSPPLIC 195 202 /FPVEKRP->VSGMGEMG (in isoform 3).
FT VARSPPLIC 203 1210 Missing (in isoform 3).
FT CONFLICT 5 5 A -> Q (IN REF. 5; AAH18718).
FT CONFLICT 13 13 A -> G (IN REF. 5; AAH20970).
FT CONFLICT 55 55 N -> T (IN REF. 1 AND 4).
FT CONFLICT 168 169 P -> TR (IN REF. 2; AAH02686).
FT CONFLICT 178 178 P -> S (IN REF. 3).
FT CONFLICT 985 985 C -> R (IN REF. 3).
FT CONFLICT 994 994 C -> R (IN REF. 6).
SQ SEQUENCE 1210 AA; 132383 MW; E8508923CA66911A CRC64;

Query Match 100.0%; Score 29; DB 1; Length 1210;
Best Local Similarity 100.0%; Pred. No. 1.7e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SSGPSL 6

605 SSGPSL 610

Db

Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C., Rodriguez A.C., Greenwood J., Schmutz J., Myers R.M., Butterfield A.S., Krzywicki M.L., Skalska U., Smalusz D.E., Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.; "Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences." Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002). [6]. SEQUENCE OF 195-1210 FROM N.A. AND CHARACTERIZATION. TISSUE=Histocytic lymphoma; PUBMED=8457211; MEDLINE=93207535; PUBMED=8457211; Milner C.M., Campbell R.D.; "The G9a gene in the human major histocompatibility complex encodes a novel protein containing ankryrin-like repeats." Biochem. J. 290:811-818 (1993). [7]. ENZYMIC ACTIVITY, AND SUBCELLULAR LOCATION. MEDLINE=21326082; PUBMED=1316813; Tachibana M., Sugimoto K., Fukushima T., Shinkai Y.; "Set domain-containing protein, G9a, is a novel lysine-preferring mammalian histone methyltransferase with hyperactivity and specific selectivity to lysines 9 and 27 of histone H3." J. Biol. Chem. 276:25309-25317 (2001). [8]. IDENTIFICATION IN COMPLEX WITH E2F6; TFDPI; MAX; MGA; EHM7ASEL; CBX3; RING1; RNF2; MBLR; L3MBTL2 AND YAF2. MEDLINE=21999559; PUBMED=12004135; Ogawa H., Ishiguro K.-I., Gaubatz S., Livingston D.M., Nakatani Y.; "A complex with chromatin modifiers that occupies E2F- and Myc-responsive genes in G0 cells." Science 296:1132-1136 (2002). -1- FUNCTION: Histone methyltransferase. Preferentially methylates Lys-9 of histone H3 and Lys-27 of histone H3 (in vitro). H3 Lys-9 methylation represents a specific tag for epigenetic transcriptional repression by recruiting HPI proteins to methylated histones. Probably targeted to histone H3 by different DNA-binding proteins like E2F6, MGA, MAX and/or Dpi. Also methylates histone H1 (By similarity). -1- CATALYTIC ACTIVITY: S-adenosyl-L-methionine + histone L-lysine = S-adenosyl-L-homocysteine + histone N(6)-methyl-L-lysine. -1- SUBUNIT: Part of the E2F6-complex in G0 phase composed of E2F6, MGA, MAX, TFDPI, CBX3, BAT8, EHM7ASEL, RING1, RNF2, MBLR, L3MBTL2 and YAF2. -1- SUBCELLULAR LOCATION: Nuclear; associates with euchromatic regions. Does not associate with heterochromatin. -1- ALTERNATIVE PRODUCTS: Event=Alternative splicing; Named isoforms=3; Comment=Additional isoforms seem to exist; Name=1; IsoId=Q96KQ7-1; Sequence=Displayed; Name=2; Synonyms=NG36G9a-SPI; IsoId=Q96KQ7-2; Sequence=VSP_002211; Name=3; Synonyms=NG36; IsoId=Q96KQ7-3; Sequence=VSP_002213; TISSUE SPECIFICITY: Expressed in all tissues examined, with high levels in fetal liver, thymus, lymph node, spleen and peripheral blood leukocytes and lower level in bone marrow. -1- SIMILARITY: Belongs to the histone-lysine methyltransferase family. Suvar3-9 subfamily. -1- SIMILARITY: Contains 7 ANK repeats. -1- SIMILARITY: Contains 1 pre-SET domain. -1- SIMILARITY: Contains 1 SET domain. -1- SIMILARITY: Contains 1 post-SET domain. -1- CAUTION: NG36 and G9a were originally thought to derive from two separate genes. Ref.3 shows that all G9a transcripts also contain the in frame coding sequence of NG36. -1- CAUTION: Ref.1 and Ref.2 sequences differ from that shown due to erroneous gene model prediction. -1- CAUTION: It is not known whether Met-1 or Met-21 is the initiator methionine.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation -

RESULT 11
 ID BAT8_MOUSE STANDARD; PRT; 1263 AA.
 AC Q92149; Q8K4R6; Q8K4R7; Q92149;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Histone-lysine N-methyltransferase, H3 lysine-9 specific 3
 DE (EC 2.1.1.43) (Histone H3-K9 methyltransferase 3) (H3-K9-HMTase 3)
 DE (HLA-B associated transcript 8) (G9a) (NG36).
 GN BAT8 OR G9A OR NG36.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=129;
 RA Rowen L., Madan A., Qin S., Hall J., Dahl T., James R., Dickhoff R.,
 RA Schaffer T., Ratcliffe A., Abbas N., Loretz C., Laskey S., Hood L.,
 RA "Sequence of the mouse MHC class III region."
 RL Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2), AND MUTAGENESIS OF
 RP 1165-ASN--CYS-1168.
 RX MEDLINE=22123403; PubMed=12130538;
 RA Tachibana M., Sugimoto K., Nozaki M., Ueda J., Ohta T., Ohki M.,
 RA Fukuda M., Takeda N., Niida H., Kato H., Shinkai Y.,
 RT "G9a histone methyltransferase plays a dominant role in euchromatic
 RT histone H3 lysine 9 methylation and is essential for early
 RT embryogenesis.";
 RL Genes Dev. 16:1779-1791 (2002).
 RN [3]
 RP SEQUENCE OF 960-1263 FROM N.A.
 RC TISSUE=Brain tumor;
 RX MEDLINE=22386257; PubMed=12477932;
 RA Strausberg R.D., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Haiech F.,
 RA Stachenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Usain T.B., Toshiyuki S., Carninci P., Prange C.,
 RA Rana S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullavy S.J.,
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko V., Bouffard G.G.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grinstead J., Schmutz J., Myers R.M.,
 RA Butterfield Y.S.N., Krzyzinski M.I., Skalska U., Smallos D.B.,
 RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.,
 RT "Generation and initial analysis of more than 15,000 full-length
 RT human and mouse cDNA sequences.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
 RN [4]
 RP ALTERNATIVE SPLICING (ISOFORM 2).
 RX MEDLINE=21564388; PubMed=1170778;
 RA Brown S.E., Campbell R.D., Sanderson C.M.;
 RT "Novel NG36/G9a gene products encoded within the human and mouse MHC
 RT class III regions.";
 RL Mamm. Genome 12:916-924 (2001).
 RN [5]
 RP ENZYMIC ACTIVITY, SUBCELLULAR LOCATION, AND MUTAGENESIS OF ARG-1162.
 RX MEDLINE=21326082; PubMed=11316813;
 RA Tachibana M., Sugimoto K., Fukushima T., Shinkai Y.;
 RT "Set domain-containing protein, G9a, is a novel lysine-preferring
 RT mammalian histone methyltransferase with hyperactivity and specific
 RT selectivity to lysines 9 and 27 of histone H3.";
 RL J. Biol. Chem. 276:25309-25317 (2001).
 RN [6]
 RP FUNCTION: Histone methyltransferase. Preferentially methylates
 CC Lys-9 of histone H3 and Lys-27 of histone H3 (in vitro). H3 Lys-9
 CC methylation represents a specific tag for epigenetic
 CC transcriptional repression by recruiting HP1 proteins to
 CC methylated histones. Also methylates histone H1.
 CC -!- CATALYTIC ACTIVITY: S-adenosyl-L-methionine + histone L-lysine =
 CC S-adenosyl-L-homocysteine + histone N(6)-methyl-L-lysine.
 CC -!- SUBUNIT: Part of the E2F6.com-1 complex in G0 phase composed of
 CC E2F6, MGA, MAX, TDP1, CBX3, BAT8, EHM1, RING1, RNF2, MBLR,
 CC L3MBTL2 and YAF2 (By similarity).
 CC -!- SUBCELLULAR LOCATION: Nuclear; almost excluded from nucleoli.
 CC Associates with euchromatic regions. Does not associate with
 CC heterochromatin.
 CC -!- ALTERNATIVE PRODUCTS:
 CC Event-Alternative splicing; Named isoforms=2;
 CC Name=1; Synonyms=G9a-L;
 CC IsoId=Q92148-1; Sequence=Displayed;
 CC Name=2; Synonyms=G9a-S;
 CC IsoId=Q92148-2; Sequence=VSP_002215, VSP_002216;
 CC -!- MISCELLANEOUS: G9a deficient mice show a higher level of histone
 CC H3 with acetylated Lys-9 and/or methylated Lys-4, display severe
 CC developmental defects and die within E9.5-E12.5 stages.
 CC -!- SIMILARITY: Belongs to the histone-lysine methyltransferase
 CC family. Suvar3-9 subfamily.
 CC -!- SIMILARITY: Contains 7 ANK repeats.
 CC -!- SIMILARITY: Contains 1 pre-SET domain.
 CC -!- SIMILARITY: Contains 1 SET domain.
 CC -!- SIMILARITY: Contains 1 post-SET domain.
 CC -!- CAUTION: NG36 and G9a were originally thought to derive from two
 CC separate genes.
 CC -!- CAUTION: Ref.1 sequences differ from that shown due to erroneous
 CC gene model prediction.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 CC EMBL; AF109906; AAC84164.1; ALT_SEQ.
 CC EMBL; AF109906; AAC84165.1; ALT_SEQ.
 CC EMBL; AB077209; BAC05482.1;
 CC EMBL; AB077210; BAC05483.1;
 CC EMBL; BC025539; AAH25539.1; ALT_INIT.
 CC MGD; MG1:2148922; Bat8.
 CC InterPro; IPR002110; ANK.
 CC InterPro; IPR003616; PostSET.
 CC InterPro; IPR007728; Pre-SET.
 CC InterPro; IPR001214; SET.
 CC InterPro; IPR003606; Zn2-binding.
 CC Pfam; PF00023; ank; 6.
 CC Pfam; PF05033; Pre-SET; 1.
 CC Pfam; PF00856; SET; 1.
 CC PRINTS; PR01415; ANKYRIN.
 CC SMART; SM00468; PreSET; 1.
 CC SMART; SM00317; SET; 1.
 CC PROSITE; PS00088; ANK_REPEAT; 5.
 CC PROSITE; PS0297; ANK_REPEAT_REGION; 1.
 CC PROSITE; PS00868; POST_SET; FALSE_NEG.
 CC PROSITE; PS00867; PRE_SET; 1.
 CC PROSITE; PS0280; SET; 1.
 CC TRANSFERASE; Methyltransferase; Chromatin regulator; Nuclear protein;
 KW ANK repeat; Repeat; Alternative splicing.
 FT DOMAIN 352..379 POLY-GLU.
 FT REPEAT 702..731 ANK 1.
 FT REPEAT 737..766 ANK 2.
 FT REPEAT 770..799 ANK 3.
 FT REPEAT 803..833 ANK 4.
 FT REPEAT 837..866 ANK 5.
 FT REPEAT 870..899 ANK 6.
 FT REPEAT 903..932 ANK 7.

FT DOMAIN 1025 1088 PRE-SET.
 FT DOMAIN 1090 1212 SET.
 FT DOMAIN 1217 1233 POST-SET.
 FT VARSPLIC 1 57 Missing (in isoform 2).
 FT VARSPLIC 58 71 /FTID=VSP 002214.
 FT VARSPLIC 426 459 /FTID=VSP 002215.
 FT VARSPLIC 1162 1162 Missing (in isoform 2).
 FT MUTAGEN 1162 1162 /FTID=VSP 002216.
 FT MUTAGEN 1165 1168 R->H: STRONGLY REDUCES HISTONE
 METHYLTRANSFERASE ACTIVITY.
 FT METHYLTRANSFERASE ACTIVITY AND SUBSEQUENT
 REPRESSION
 SQ SEQUENCE 1263 AA; 138038 MW; 74DEFF9A36769589 CRC64;
 Query Match 100.0%; Score 29; DB 1; Length 1263;
 Best Local Similarity 100.0%; Pred. No. 1.8e+02;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 SSGPSL 6
 DB 658 SSGPSL 663

RESULT 12
 ZMSL_YEAST STANDARD; PRT; 1380 AA.
 AC P46974;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE Zinc finger protein ZMS1.
 GN ZMS1 OR YJR127C OR J2052.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
 OX NCBI_TaxID=4932;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=RC11-6A;
 RA Thomas D., Barbey R., Surdin-Kerjan Y.;
 RL Submitted (XX-1993) to the EMBL/GenBank/DBJ databases.
 [2]
 RP SEQUENCE FROM N.A.
 RA Rose M., Koetter P., Entian K.D.;
 RL Submitted (SEP-1993) to the EMBL/GenBank/DBJ databases.
 CC -1- SUBCELLULAR LOCATION: Nuclear (Potential).
 CC -1- SIMILARITY: Contains 2 C2H2-type zinc fingers.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sb-sib.ch).
 CC -----
 CC EMBL; L26506; AAA3240.1; --
 CC EMBL; Z49627; CAA89658.1; --
 CC PIR; S57150; S57150.
 CC HSSP; P07248; 1ARD.
 CC GerMOnline; 141960; --
 CC TRANSFAC; T04595; --
 CC SGD; S0003888; ZMS1.
 CC InterPro; IPR007087; Znf_C2H2.
 CC Pfam; PF00096; Zf-C2H2_2.
 CC SMART; SM00355; Znf_C2H2; 2.
 CC PROSITE; PS00028; ZINC_FINGER_C2H2_1; 2.
 CC PROSITE; PS0157; ZINC_FINGER_C2H2_2; 2.
 CC DNA-binding; Nuclear protein; Zinc-finger; Metal-binding; Repeat.
 CC DOMAIN 64 67 POLY-PHE.
 FT

FT ZN_FING 151 173 C2H2-TYPE 1.
 FT ZN_FING 179 202 C2H2-TYPE 2.
 FT DOMAIN 444 433 POLY-GLN
 FT CONFLICT 1 17 MEPPAFGAPALCILT -> MHTN (IN REF. 1).
 FT CONFLICT 1117 1118 NL -> IF (IN REF. 1).
 FT CONFLICT 1131 1131 N -> H (IN REF. 1).
 FT CONFLICT 1143 1380 RLIIPVLFKMRRCCLDAHVIEKTLWKDWSNMKALBEVC
 YDMGSLREATEYALNNVDATWTFYIKGGRIFNTVFA
 TTMFTAVLIVSEYKNCVEDWARGVANNPNSALLDFSDRV
 LMKAEIRLRRQMLIPKCECDLVKSYTDFLRWQDKDALDL
 SALNEEQQRADPNPTDINETIQLIVAAASLSKCLYLGVQI
 LGDAIWPILLSFAHGLQSRALYSVTKKENTRI -> S
 (IN REF. 1).
 SQ SEQUENCE 1380 AA; 155062 MW; FEE987EA5B639000 CRC64;
 Query Match 100.0%; Score 29; DB 1; Length 1380;
 Best Local Similarity 100.0%; Pred. No. 1.9e+02;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 SSGPSL 6
 DB 458 SSGPSL 463

RESULT 13
 SPEN_DROME STANDARD; PRT; 5560 AA.
 AC Q8SX83; Q9NH17; Q9UC63; Q9VPL1; Q9VPL2;
 DT 10-OCT-2003 (Rel. 42, Created)
 DT 10-OCT-2003 (Rel. 42, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Split ends protein.
 GN SPEN OR CGL8497.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORMS 3 AND 4). FUNCTION, ALTERNATIVE PROMOTER
 RP USAGE, TISSUE SPECIFICITY, DEVELOPMENTAL STAGE, AND MUTANTS E9 AND
 RP D57.
 RC TISSUE=Embryo;
 RX MEDLINE=20025936; PubMed=10556062;
 RA Wellelette E.L., Harding K.W., Mace K.A., Ronshaugen M.R., Wang F.Y.,
 RA McGinnis W.;
 RA "spen encodes an RNP motif protein that interacts with Hox pathways
 RT to repress the development of head-like sclerites in the Drosophila
 RT trunk";
 RL Development 126:5373-5385 (1999).
 RN [2]
 RP SEQUENCE FROM N.A. (ISOFORM 4).
 RC TISSUE=Embryo;
 RX MEDLINE=20157049; PubMed=10655223;
 RA Rebay L., Chen F., Hsiao F., Kolodziej P.A., Kuang B.H., Laverly T.,
 RA Suh C., Voas M., Williams A., Rubin G.M.;
 RA "A genetic screen for novel components of the Ras/mitogen-activated
 RT protein kinase signaling pathway that interact with the Yan gene of
 RT Drosophila identifies split ends, a new RNA recognition motif-
 RT containing protein";
 RL Genetics 154:695-712 (2000).
 RN [3]
 RP SEQUENCE FROM N.A. (ISOFORM 1), SUBCELLULAR LOCATION, AND TISSUE
 RP SPECIFICITY.
 RC TISSUE=Embryo;
 RX MEDLINE=20171275; PubMed=10704397;
 RA Kuang B.H., Wu S.C.-Y., Shin Y.-A., Luo L., Kolodziej P.A.;
 RA "split ends encodes large nuclear proteins that regulate neuronal
 RT cell fate and axon extension in the Drosophila embryo";
 RL Development 127:1517-1529 (2000).
 RN [4]
 RP SEQUENCE FROM N.A., AND TISSUE SPECIFICITY.
 RC STRAIN=Berkley;

RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scher S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Vandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA April J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
 RA Bailew K.Y., Basu A., Baxendale J., Bayraktaroglu L., Beaskey E.M.,
 RA Beeson K.Y., Benson P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borokova D., Botchan M.R., Bouck J., Brockstein P., Brottier P.,
 RA Burris K.C., Busan D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferrieria S., Fleischmann W.,
 RA Fiesler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D.A., Heinman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasko P., Lei Y., Levitsky A.A., Li J.H., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusser D.R., Pacleb J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler P., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Stimpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin C.M., Venter J.C.;
 RA "The genome sequence of *Drosophila melanogaster*.";
 RL Science 287:2185-2195(2000).
 RN [5]
 RP REVISIONS, AND ALTERNATIVE SPLICING.
 RX MEDLINE=22426069; PubMed=12537572;
 RA Mitra S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,
 RA Hradecky P., Huang Y., Kaminker J.S., Millburn G.H., Prochuk S.E.,
 RA Smith C.D., Tupy J.L., Whitfield E.J., Bayraktaroglu L., Berman B.P.,
 RA Bettencourt B.R., Celniker S.E., de Grey A.D.N.J., Drysdale R.A.,
 RA Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.Q.,
 RA Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,
 RA Lewis S.E.;
 RA "Annotation of the *Drosophila melanogaster* euchromatic genome: a
 RT systematic review.";
 RL Genome Biol. 3:RESEARCH0083.1-RESEARCH0083.22(2002).
 RN [6]
 RP SEQUENCE OF 424-2002 FROM N.A.
 RC STRAIN=Berkeley; TISSUE=Embryo;
 RX MEDLINE=22426066; PubMed=12537569;
 RA Stapleton M., Carlson J.W., Brokstein P., Yu C., Champe M.,
 RA George R.A., Guarin H., Krommiller B., Pacleb J.M., Park S., Wan K.H.,
 RA Rubin G.M., Celniker S.E.;
 RA "A *Drosophila* full-length cDNA resource.";
 RL Genome Biol. 3:RESEARCH0080.1-RESEARCH0080.8(2002).
 RN [7]
 RP FUNCTION.
 RX MEDLINE=20253107; PubMed=10790398;
 RA Lane M.E., Elend M., Heidmann D., Herr A., Marzodko S., Herrig A.,
 RA Lehner C.F.;
 RT "A screen for modifiers of cyclin E function in *Drosophila*
 RT melanogaster identifies Cdk2 mutations, revealing the insignificance
 RT of putative phosphorylation sites in Cdk2.";
 RL Genetics 155:233-244(2000).
 RN [8]
 RP FUNCTION ON EGF RECEPTOR PATHWAY.
 RX MEDLINE=20414403; PubMed=10959845;
 RA Chen F., Rebay I.;
 RT "split ends, a new component of the *Drosophila* EGF receptor pathway,
 RL regulates development of midline glial cells.";
 RL Curr. Biol. 10:943-946(2000).
 RN [9]
 RP FUNCTION ON WG PATHWAY.
 RX MEDLINE=22668876; PubMed=12783785;
 RA Lin H.V., Dorquez D.B., Cho S., Chen F., Rebay I., Cadigan K.M.;
 RT "Split ends is a tissue/promoter specific regulator of Wingless
 RT signaling.";
 RL Development 130:3125-3135(2003).
 CC -I- FUNCTION: Probable corepressor protein, which regulates different
 CC key pathways such as the EGF receptor and Wg pathways. Involved in
 CC neuronal cell fate, survival and axon guidance, cell cycle
 CC regulation and repression of head identity in the embryonic trunk.
 CC May act with the Hox gene *Deformed* and the EGF receptor signaling
 CC pathway. Positive regulator of the Wg pathway in larval tissues
 CC but not in embryonic tissues. May act as a transcriptional
 CC corepressor protein, which repress transcription via the
 CC recruitment of large complexes containing histone deacetylase
 CC proteins.
 CC -I- SUBCELLULAR LOCATION: Nuclear.
 CC -I- ALTERNATIVE PRODUCTS:
 CC Event=Alternative promoter;
 CC Comment=2 isoforms, 1 (shown here) and 2, are produced by use of
 CC Event=Alternative promoters;
 CC Event=Alternative splicing; Named isoforms=4;
 CC Name=1;
 CC IsoId=Q8X83-1; Sequence=Displayed;
 CC Name=2;
 CC IsoId=Q8X83-2; Sequence=VSP_008565, VSP_008566;
 CC Note=No experimental confirmation available;
 CC Name=3; Synonyms=SpnL;
 CC IsoId=Q8X83-3; Sequence=VSP_008567;
 CC Note=Produced by alternative splicing of isoform 1;
 CC Name=4; Synonyms=SpnS;
 CC IsoId=Q8X83-4; Sequence=VSP_008565, VSP_008566, VSP_008567;
 CC Note=Produced by alternative splicing of isoform 2;
 CC -I- TISSUE SPECIFICITY: Ubiquitous. Expressed prior to cellularization
 CC in stage 3 embryos, and in blastoderm cells, including pole cells.
 CC Expressed throughout the rest of embryogenesis. Later, it is
 CC expressed at higher level in epidermal cells and CNS.
 CC -I- DEVELOPMENTAL STAGE: Isoform 3 is expressed both maternally and
 CC zygotically.
 CC -I- SIMILARITY: Belongs to the Spn family.
 CC -I- SIMILARITY: Contains 3 RNA recognition motif (RRM) domains.
 CC -I- SIMILARITY: Contains 1 SPOC domain.
 CC -I- CAUTION: It is uncertain whether Met-1 or Met-7 is the initiator.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; AF188205; AAF13218.1; -;
 CC EMBL; AF184612; AAF26299.1; -;
 CC EMBL; AF221715; AAF34561.1; ALT_INIT.
 CC EMBL; AE003590; AAF51334.2; -;
 CC EMBL; AE003590; AAF51335.2; -;
 CC EMBL; AE003590; AAF10511.1; -;
 CC EMBL; AY094788; AAM11141.1; ALT_SEQ.
 CC HSP; P09651; IHA1.
 CC FlyBase: FBgn0016977; spen.
 CC GO; GO:0007411; P:axon guidance; IMP.
 CC GO; GO:0008447; P:glia cell migration; IMP.
 CC InterPro: IPR000504; RNA_rec_mot.
 CC Pfam; PF00076; rim; 3.
 CC SMART; SM00360; RRM; 3.
 CC PROSITE; PS0102; RRM; 3.
 CC PROSITE; PS00030; RRM_RNP_1; FALSE_NEG.

DR PROSITE; PS00917; SPOC; 1.
 KW Transcription regulation; Repressor; Developmental protein;
 KW Nuclear protein; Repeat; RNA-binding; Coiled coil;
 KW Alternative promoter usage; Alternative splicing.
 FT DOMAIN 554 632 RNA-BINDING (RRM) 1.
 FT DOMAIN 656 730 RNA-BINDING (RRM) 2.
 FT DOMAIN 734 806 RNA-BINDING (RRM) 3.

Query Match 100.0%; Score 29; DB 1; Length 5560;
 Best Local Similarity 100.0%; Pred. No. 8.3e+02; Indels 0; Gaps 0;
 Matches 6; Conservative 0; Mismatches 0;

QY 1 SSGPSL 6
 DB 3430 SSGPSL 3435

RESULT 14
 KCH2_CHICK STANDARD; PRT; 526 AA.
 ID Q9PT84;
 AC 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 15-MAR-2004 (Rel. 43, Last annotation update)
 DE Potassium voltage-gated channel subfamily H member 2 (Ether-a-go-go
 DE related gene potassium channel 1) (ERG1) (Ether-a-go-go related
 DE protein 1) (Erg related protein 1) (Fragment).
 GN KCNH2 OR ERG.
 OS Gallus gallus (Chicken).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
 OC Gallus.
 OC NCBI_TaxID=9031;
 [1]
 RN SEQUENCE FROM N.A.
 RC TISSUE=Embryo;
 RX MEDLINE=20368176; PubMed=10906470;
 RA Crociani O., Cherubini G., Rosati B., Wanke E., Olivetto M.,
 RA Arcangeli A., Hofmann G., Rossi B., Wanke E., Olivetto M.,
 RA Arcangeli A., Hofmann G., Rossi B., Wanke E., Olivetto M.,
 RT "erg gene(s) expression during development of the nervous and muscular
 RT system of quail embryos.";
 RL Mech. Dev. 95:239-243(2000).
 CC -|- FUNCTION: Pore-forming (alpha) subunit of voltage-gated inwardly
 CC rectifying potassium channel (By similarity).
 CC -|- SUBUNIT: The potassium channel is probably composed of a homo- or
 CC heterotetrameric complex of pore-forming alpha subunits that can
 CC associate with modulating beta subunits (By similarity).
 CC -|- SUBCELLULAR LOCATION: Integral membrane protein.
 CC -|- DOMAIN: The segment S4 is probably the voltage-sensor and is
 CC characterized by a series of positively charged amino acids at
 CC every third position.
 CC -|- SIMILARITY: Belongs to the potassium channel family. H (Erg)
 CC subfamily.
 CC -|- SIMILARITY: Contains 1 cyclic nucleotide-binding domain.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.ebi.ac.uk/announcements>
 CC or send an email to license@ebi.ac.uk).

DR EMBL; AJ271210; CAB66135.1; -
 DR InterPro; IPR000595; CNMP_binding.
 DR InterPro; IPR003967; Erg_channel.
 DR InterPro; IPR003821; Ion_trans.
 DR InterPro; IPR001622; K-channel_pore.
 DR InterPro; IPR005820; M-channel_flg.
 DR Pfam; PF00027; CNMP_trans; 1.
 DR Pfam; PF00520; Ion_trans; 1.
 DR PRINTS; PR01470; ERGCHANNEL.

DR SMART; SM00100; CNMP; 1.
 DR PROSITE; PS00888; CNMP_BINDING_1; FALSE_NEG.
 DR PROSITE; PS00889; CNMP_BINDING_2; FALSE_NEG.
 DR PROSITE; PS00042; CNMP_BINDING_3; 1.
 KW Transport; Ion transport; Ionic channel; Voltage-gated channel;
 KW Potassium channel; Potassium; Potassium transport; Transmembrane;
 KW Glycoprotein; Multigene family.
 FT DOMAIN 1 1
 FT NON_TER 1
 FT DOMAIN 103 103 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 104 124 SEGMENT S1 (POTENTIAL).
 FT TRANSMEM 151 171 SEGMENT S2 (POTENTIAL).
 FT DOMAIN 172 192 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 193 213 SEGMENT S3 (POTENTIAL).
 FT TRANSMEM 224 244 SEGMENT S4 (POTENTIAL).
 FT DOMAIN 245 250 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 251 271 SEGMENT S5 (POTENTIAL).
 FT TRANSMEM 316 336 SEGMENT H5 (PORE-FORMING) (POTENTIAL).
 FT TRANSMEM 343 363 SEGMENT S6 (POTENTIAL).
 FT DOMAIN 364 526 CYTOPLASMIC (POTENTIAL).
 FT NP_BIND 446 >526 CNMP.
 FT CARBOHYD 144 144 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 301 301 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT NON_TER 526
 SQ SEQUENCE 526 AA; 59793 MW; 2C4D6DBD1A9E015C CRC64;

Query Match 93.1%; Score 27; DB 1; Length 526;
 Best Local Similarity 83.3%; Pred. No. 2e+02; Indels 0; Gaps 0;
 Matches 5; Conservative 1; Mismatches 0;

QY 1 SSGPSL 6
 DB 306 SSGPSI 311

RESULT 15
 MYCD_MOUSE STANDARD; PRT; 935 AA.
 ID MYCD_MOUSE
 AC Q8VIMS; Q8C3W6; Q8VIL4;
 DT 10-OCT-2003 (Rel. 42, Created)
 DT 10-OCT-2003 (Rel. 42, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Myocardin (SRF co-factor protein) (Basic SAP coiled-coil transcription
 DE activator 2).
 GN MyCD OR SRF OR BSAC2.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OC NCBI_TaxID=10090;
 [1]
 RN SEQUENCE FROM N.A. FUNCTION, TISSUE SPECIFICITY, DEVELOPMENTAL STAGE,
 RN SUBCELLULAR LOCATION, INTERACTION WITH SRF, AND MUTAGENESIS OF
 RN 387-GLU--ARG-389 AND 408-ASP--LEU-410.
 RC TISSUE=Heart;
 RX MEDLINE=21332583; PubMed=11439182;
 RX Wang D.-Z., Chang P.S., Wang Z., Sutherland L., Richardson J.A.,
 RA Small E., Krieg P.A., Olson E.N.;
 RA "Activation of cardiac gene expression by myocardin, a transcriptional
 RA cofactor for serum response factor.";
 RL Cell 105:851-862(2001).
 [2]
 RN SEQUENCE FROM N.A. AND REVISIONS OF 1-128.
 RN MEDLINE=22317395; PubMed=12397177;
 RX Wang D.-Z., Li S., Hockemeyer D., Sutherland L., Wang Z., Schmitt G.,
 RA Richardson J.A., Nordheim A., Olson E.N.;
 RA "Potentiation of serum response factor activity by a family of
 RA myocardin-related transcription factors";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:14855-14860(2002).
 [3]
 RN SEQUENCE FROM N.A. (ISOFORM 2).
 RN STRAIN=C57BL/6; TISSUE=Heart;
 RA Sawada T., Okazaki T., Nakano H.;
 RA "An alternative splicing form of myocardin (BSAC2), myocardin A
 RA (BSAC2A).";

RL
RN
RP
RX
RY
RZ
S1
S2
S3
S4
S5
S6
S7
S8
S9
S10
S11
S12
S13
S14
S15
S16
S17
S18
S19
S20
S21
S22
S23
S24
S25
S26
S27
S28
S29
S30
S31
S32
S33
S34
S35
S36
S37
S38
S39
S40
S41
S42
S43
S44
S45
S46
S47
S48
S49
S50
S51
S52
S53
S54
S55
S56
S57
S58
S59
S60
S61
S62
S63
S64
S65
S66
S67
S68
S69
S70
S71
S72
S73
S74
S75
S76
S77
S78
S79
S80
S81
S82
S83
S84
S85
S86
S87
S88
S89
S90
S91
S92
S93
S94
S95
S96
S97
S98
S99
S100
S101
S102
S103
S104
S105
S106
S107
S108
S109
S110
S111
S112
S113
S114
S115
S116
S117
S118
S119
S120
S121
S122
S123
S124
S125
S126
S127
S128
S129
S130
S131
S132
S133
S134
S135
S136
S137
S138
S139
S140
S141
S142
S143
S144
S145
S146
S147
S148
S149
S150
S151
S152
S153
S154
S155
S156
S157
S158
S159
S160
S161
S162
S163
S164
S165
S166
S167
S168
S169
S170
S171
S172
S173
S174
S175
S176
S177
S178
S179
S180
S181
S182
S183
S184
S185
S186
S187
S188
S189
S190
S191
S192
S193
S194
S195
S196
S197
S198
S199
S200
S201
S202
S203
S204
S205
S206
S207
S208
S209
S210
S211
S212
S213
S214
S215
S216
S217
S218
S219
S220
S221
S222
S223
S224
S225
S226
S227
S228
S229
S230
S231
S232
S233
S234
S235
S236
S237
S238
S239
S240
S241
S242
S243
S244
S245
S246
S247
S248
S249
S250
S251
S252
S253
S254
S255
S256
S257
S258
S259
S260
S261
S262
S263
S264
S265
S266
S267
S268
S269
S270
S271
S272
S273
S274
S275
S276
S277
S278
S279
S280
S281
S282
S283
S284
S285
S286
S287
S288
S289
S290
S291
S292
S293
S294
S295
S296
S297
S298
S299
S300
S301
S302
S303
S304
S305
S306
S307
S308
S309
S310
S311
S312
S313
S314
S315
S316
S317
S318
S319
S320
S321
S322
S323
S324
S325
S326
S327
S328
S329
S330
S331
S332
S333
S334
S335
S336
S337
S338
S339
S340
S341
S342
S343
S344
S345
S346
S347
S348
S349
S350
S351
S352
S353
S354
S355
S356
S357
S358
S359
S360
S361
S362
S363
S364
S365
S366
S367
S368
S369
S370
S371
S372
S373
S374
S375
S376
S377
S378
S379
S380
S381
S382
S383
S384
S385
S386
S387
S388
S389
S390
S391
S392
S393
S394
S395
S396
S397
S398
S399
S400
S401
S402
S403
S404
S405
S406
S407
S408
S409
S410
S411
S412
S413
S414
S415
S416
S417
S418
S419
S420
S421
S422
S423
S424
S425
S426
S427
S428
S429
S430
S431
S432
S433
S434
S435
S436
S437
S438
S439
S440
S441
S442
S443
S444
S445
S446
S447
S448
S449
S450
S451
S452
S453
S454
S455
S456
S457
S458
S459
S460
S461
S462
S463
S464
S465
S466
S467
S468
S469
S470
S471
S472
S473
S474
S475
S476
S477
S478
S479
S480
S481
S482
S483
S484
S485
S486
S487
S488
S489
S490
S491
S492
S493
S494
S495
S496
S497
S498
S499
S500
S501
S502
S503
S504
S505
S506
S507
S508
S509
S510
S511
S512
S513
S514
S515
S516
S517
S518
S519
S520
S521
S522
S523
S524
S525
S526
S527
S528
S529
S530
S531
S532
S533
S534
S535
S536
S537
S538
S539
S540
S541
S542
S543
S544
S545
S546
S547
S548
S549
S550
S551
S552
S553
S554
S555
S556
S557
S558
S559
S560
S561
S562
S563
S564
S565
S566
S567
S568
S569
S570
S571
S572
S573
S574
S575
S576
S577
S578
S579
S580
S581
S582
S583
S584
S585
S586
S587
S588
S589
S590
S591
S592
S593
S594
S595
S596
S597
S598
S599
S600
S601
S602
S603
S604
S605
S606
S607
S608
S609
S610
S611
S612
S613
S614
S615
S616
S617
S618
S619
S620
S621
S622
S623
S624
S625
S626
S627
S628
S629
S630
S631
S632
S633
S634
S635
S636
S637
S638
S639
S640
S641
S642
S643
S644
S645
S646
S647
S648
S649
S650
S651
S652
S653
S654
S655
S656
S657
S658
S659
S660
S661
S662
S663
S664
S665
S666
S667
S668
S669
S670
S671
S672
S673
S674
S675
S676
S677
S678
S679
S680
S681
S682
S683
S684
S685
S686
S687
S688
S689
S690
S691
S692
S693
S694
S695
S696
S697
S698
S699
S700
S701
S702
S703
S704
S705
S706
S707
S708
S709
S710
S711
S712
S713
S714
S715
S716
S717
S718
S719
S720
S721
S722
S723
S724
S725
S726
S727
S728
S729
S730
S731
S732
S733
S734
S735
S736
S737
S738
S739
S740
S741
S742
S743
S744
S745
S746
S747
S748
S749
S750
S751
S752
S753
S754
S755
S756
S757
S758
S759
S760
S761
S762
S763
S764
S765
S766
S767
S768
S769
S770
S771
S772
S773
S774
S775
S776
S777
S778
S779
S780
S781
S782
S783
S784
S785
S786
S787
S788
S789
S790
S791
S792
S793
S794
S795
S796
S797
S798
S799
S800
S801
S802
S803
S804
S805
S806
S807
S808
S809
S810
S811
S812
S813
S814
S815
S816
S817
S818
S819
S820
S821
S822
S823
S824
S825
S826
S827
S828
S829
S830
S831
S832
S833
S834
S835
S836
S837
S838
S839
S840
S841
S842
S843
S844
S845
S846
S847
S848
S849
S850
S851
S852
S853
S854
S855
S856
S857
S858
S859
S860
S861
S862
S863
S864
S865
S866
S867
S868
S869
S870
S871
S872
S873
S874
S875
S876
S877
S878
S879
S880
S881
S882
S883
S884
S885
S886
S887
S888
S889
S890
S891
S892
S893
S894
S895
S896
S897
S898
S899
S900
S901
S902
S903
S904
S905
S906
S907
S908
S909
S910
S911
S912
S913
S914
S915
S916
S917
S918
S919
S920
S921
S922
S923
S924
S925
S926
S927
S928
S929
S930
S931
S932
S933
S934
S935
S936
S937
S938
S939
S940
S941
S942
S943
S944
S945
S946
S947
S948
S949
S950
S951
S952
S953
S954
S955
S956
S957
S958
S959
S960
S961
S962
S963
S964
S965
S966
S967
S968
S969
S970
S971
S972
S973
S974
S975
S976
S977
S978
S979
S980
S981
S982
S983
S984
S985
S986
S987
S988
S989
S990
S991
S992
S993
S994
S995
S996
S997
S998
S999
S1000
S1001
S1002
S1003
S1004
S1005
S1006
S1007
S1008
S1009
S1010
S1011
S1012
S1013
S1014
S1015
S1016
S1017
S1018
S1019
S1020
S1021
S1022
S1023
S1024
S1025
S1026
S1027
S1028
S1029
S1030
S1031
S1032
S1033
S1034
S1035
S1036
S1037
S1038
S1039
S1040
S1041
S1042
S1043
S1044
S1045
S1046
S1047
S1048
S1049
S1050
S1051
S1052
S1053
S1054
S1055
S1056
S1057
S1058
S1059
S1060
S1061
S1062
S1063
S1064
S1065
S1066
S1067
S1068
S1069
S1070
S1071
S1072
S1073
S1074
S1075
S1076
S1077
S1078
S1079
S1080
S1081
S1082
S1083
S1084
S1085
S1086
S1087
S1088
S1089
S1090
S1091
S1092
S1093
S1094
S1095
S1096
S1097
S1098
S1099
S1100
S1101
S1102
S1103
S1104
S1105
S1106
S1107
S1108
S1109
S1110
S1111
S1112
S1113
S1114
S1115
S1116
S1117
S1118
S1119
S1120
S1121
S1122
S1123
S1124
S1125
S1126
S1127
S1128
S1129
S1130
S1131
S1132
S1133
S1134
S1135
S1136
S1137
S1138
S1139
S1140
S1141
S1142
S1143
S1144
S1145
S1146
S1147
S1148
S1149
S1150
S1151
S1152
S1153
S1154
S1155
S1156
S1157
S1158
S1159
S1160
S1161
S1162
S1163
S1164
S1165
S1166
S1167
S1168
S1169
S1170
S1171
S1172
S1173
S1174
S1175
S1176
S1177
S1178
S1179
S1180
S1181
S1182
S1183
S1184
S1185
S1186
S1187
S1188
S1189
S1190
S1191
S1192
S1193
S1194
S1195
S1196
S1197
S1198
S1199
S1200
S1201
S1202
S1203
S1204
S1205
S1206
S1207
S1208
S1209
S1210
S1211
S1212
S1213
S1214
S1215
S1216
S1217
S1218
S1219
S1220
S1221
S1222
S1223
S1224
S1225
S1226
S1227
S1228
S1229
S1230
S1231
S1232
S1233
S1234
S1235
S1236
S1237
S1238
S1239
S1240
S1241
S1242
S1243
S1244
S1245
S1246
S1247
S1248
S1249
S1250
S1251
S1252
S1253
S1254
S1255
S1256
S1257
S1258
S1259
S1260
S1261
S1262
S1263
S1264
S1265
S1266
S1267
S1268
S1269
S1270
S1271
S1272
S1273
S1274
S1275
S1276
S1277
S1278
S1279
S1280
S1281
S1282
S1283
S1284
S1285
S1286
S1287
S1288
S1289
S1290
S1291
S1292
S1293
S1294
S1295
S1296
S1297
S1298
S1299
S1300
S1301
S1302
S1303
S1304
S1305
S1306
S1307
S1308
S1309
S1310
S1311
S1312
S1313
S1314
S1315
S1316
S1317
S1318
S1319
S1320
S1321
S1322
S1323
S1324
S1325
S1326
S1327
S1328
S1329
S1330
S1331
S1332
S1333
S1334
S1335
S1336
S1337
S1338
S1339
S1340
S1341
S1342
S1343
S1344
S1345
S1346
S1347
S1348
S1349
S1350
S1351
S1352
S1353
S1354
S1355
S1356
S1357
S1358
S1359
S1360
S1361
S1362
S1363
S1364
S1365
S1366
S1367
S1368
S1369
S1370
S1371
S1372
S1373
S1374
S1375
S1376
S1377
S1378
S1379
S1380
S1381
S1382
S1383
S1384
S1385
S1386
S1387
S1388
S1389
S1390
S1391
S1392
S1393
S1394
S1395
S1396
S1397
S1398
S1399
S1400
S1401
S1402
S1403
S1404
S1405
S1406
S1407
S1408
S1409
S1410
S1411
S1412
S1413
S1414
S1415
S1416
S1417
S1418
S1419
S1420
S1421
S1422
S1423
S1424
S1425
S1426
S1427
S1428
S1429
S1430
S1431
S1432
S1433
S1434
S1435
S1436
S1437
S1438
S1439
S1440
S1441
S1442
S1443
S1444
S1445
S1446
S1447
S1448
S1449
S1450
S1451
S1452
S1453
S1454
S1455
S1456
S1457
S1458
S1459
S1460
S1461
S1462
S1463
S1464
S1465
S1466
S1467
S1468
S1469
S1470
S1471
S1472
S1473
S1474
S1475
S1476
S1477
S1478
S1479
S1480
S1481
S1482
S1483
S1484
S1485
S1486
S1487
S1488
S1489
S1490
S1491
S1492
S1493
S1494
S1495
S1496
S1497
S1498
S1499
S1500
S1501
S1502
S1503
S1504
S1505
S1506
S1507
S1508
S1509
S1510
S1511
S1512
S1513
S1514
S1515
S1516
S1517
S1518
S1519
S1520
S1521
S1522
S1523
S1524
S1525
S1526
S1527
S1528
S1529
S1530
S1531
S1532
S1533
S1534
S1535
S1536
S1537
S1538
S1539
S1540
S1541
S1542
S1543
S1544
S1545
S1546
S1547
S1548
S1549
S1550
S1551
S1552
S1553
S1554
S1555
S1556
S1557
S1558
S1559
S1560
S1561
S1562
S1563
S1564
S1565
S1566
S1567
S1568
S1569
S1570
S1571
S1572
S1573
S1574
S1575
S1576
S1577
S1578
S1579
S1580
S1581
S1582
S1583
S1584
S1585
S1586
S1587
S1588
S1589
S1590
S1591
S1592
S1593
S1594
S1595
S1596
S1597
S1598
S1599
S1600
S1601
S1602
S1603
S1604
S1605
S1606
S1607
S1608
S1609
S1610
S1611
S1612
S1613
S1614
S1615
S1616
S1617
S1618
S1619
S1620
S1621
S1622
S1623
S1624
S1625
S1626
S1627
S1628
S1629
S1630
S1631
S1632
S1633
S1634
S1635
S1636
S1637
S1638
S1639
S1640
S1641
S1642
S1643
S1644
S1645
S1646
S1647
S1648
S1649
S1650
S1651
S1652
S1653
S1654
S1655
S1656
S1657
S1658
S1659
S1660
S1661
S1662
S1663
S1664
S1665
S1666
S1667
S1668
S1669
S1670
S1671
S1672
S1673
S1674
S1675
S1676
S1677
S1678
S1679
S1680
S1681
S1682
S1683
S1684
S1685
S1686
S1687
S1688
S1689
S1690
S1691
S1692
S1693
S1694
S1695
S1696
S1697
S1698
S1699
S1700
S1701
S1702
S1703
S1704
S1705
S1706
S1707
S1708
S1709
S1710
S1711
S1712
S1713
S1714
S1715
S1716
S1717
S1718
S1719
S1720
S1721
S1722
S1723
S1724
S1725
S1726
S1727
S1728
S1729
S1730
S1731
S1732
S1733
S1734
S1735
S1736
S1737
S1738
S1739
S1740
S1741
S1742
S1743
S1744
S1745
S1746
S1747
S1748
S1749
S1750
S1751
S1752
S1753
S1754
S1755
S1756
S1757
S1758
S1759
S1760
S1761
S1762
S1763
S1764
S1765
S1766
S1767
S1768
S1769
S1770
S1771
S1772
S1773
S1774
S1775
S1776
S1777
S1778
S1779
S1780
S1781
S1782
S1783
S1784
S1785
S1786
S1787
S1788
S1789
S1790
S1791
S1792
S1793
S1794
S1795
S1796
S1797
S1798
S1799
S1800
S1801
S1802
S1803
S1804
S1805
S1806
S1807
S1808
S1809
S1810
S1811
S1812
S1813
S1814
S1815
S1816
S1817
S1818
S1819
S1820
S1821
S1822
S1823
S1824
S1825
S1826
S1827
S1828
S1829
S1830
S1831
S1832
S1833
S1834
S1835
S1836
S1837
S1838
S1839
S1840
S1841
S1842
S1843
S1844
S1845
S1846
S1847
S1848
S1849
S1850
S1851
S1852
S1853
S1854
S1855
S1856
S1857
S1858
S1859
S1860
S1861
S1862
S1863
S1864
S1865
S1866
S1867
S1868
S1869
S1870
S1871
S1872
S1873
S1874
S1875
S1876
S1877
S1878
S1879
S1880
S1881
S1882
S1883
S1884
S1885
S1886
S1887
S1888
S1889
S1890
S1891
S1892
S1893
S1894
S1895
S1896
S1897
S1898
S1899
S1900
S1901
S1902
S1903

FT CONFLICT 110 110 E -> EA (IN REP. 4).
FT CONFLICT 794 794 G -> D (IN REP. 3 AND 5).
SQ SEQUENCE 935 AA; 101373 MW; C1D883EF8C181149 CRC64;
Query Match 93.1%; Score 27; DB 1; Length 935;
Best Local Similarity 83.3%; Pred. No. 3.6e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 1 SSGPSL 6
DB 906 SSGPSI 911
Search completed: March 10, 2004, 09:13:52
Job time : 3.09728 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 10, 2004, 08:58:54 ; Search time 1.8677 Seconds
(without alignments)
309.015 Million cell updates/sec

Title: US-09-848-834A-6

Perfect score: 29

Sequence: 1 SSGPSL 6

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR 78.*

1: PIR1.*

2: PIR2.*

3: PIR3.*

4: PIR4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	29	100.0	225	1 JQ2040	nonstructural prot
2	29	100.0	280	2 A53027	transcription fact
3	29	100.0	348	2 AB3260	hypothetical membr
4	29	100.0	394	2 JQ2041	polymerase-associ
5	29	100.0	405	2 B97171	uncharacterized co
6	29	100.0	532	2 S74453	hypothetical prote
7	29	100.0	989	2 T46659	period clock prote
8	29	100.0	937	2 S44457	period clock prote
9	29	100.0	1001	2 S30385	G9a protein - huma
10	29	100.0	1074	2 T04777	hypothetical prote
11	29	100.0	1380	2 S57150	ZMS1 protease - yea
12	27	93.1	222	2 T34060	hypothetical prote
13	27	93.1	1487	2 T03850	hypothetical prote
14	27	93.1	1584	2 T27121	hypothetical prote
15	26	89.7	52	2 C38230	inorganic diphosph
16	26	89.7	76	2 C36817	ORF Z protein - hu
17	26	89.7	102	1 CCEG	cytochrome c [vali
18	26	89.7	102	2 H83089	conserved hypotet
19	26	89.7	137	2 H75471	hypothetical prote
20	26	89.7	156	2 H73621	hypothetical prote
21	26	89.7	209	1 HSX11A	histone H1A - Afri
22	26	89.7	229	2 I51227	histone H1A - Afri
23	26	89.7	260	2 AH2878	hypothetical prote
24	26	89.7	260	2 B97655	probable ATP-bindi
25	26	89.7	263	2 A87149	conserved hypotet
26	26	89.7	285	2 T20566	hypothetical prote
27	26	89.7	297	2 T21387	hypothetical prote
28	26	89.7	302	2 G70614	hypothetical prote
29	26	89.7	326	2 JQ3659	hypothetical 36K p

30 26 89.7 326 2 JQ1660
31 26 89.7 326 2 AC2244
32 26 89.7 340 2 G59181
33 26 89.7 366 1 A26447
34 26 89.7 393 2 B90169
35 26 89.7 399 1 S71480
36 26 89.7 457 2 S01839
37 26 89.7 460 2 H36536
38 26 89.7 464 2 T36256
39 26 89.7 470 2 C75591
40 26 89.7 491 2 T52398
41 26 89.7 500 2 T35666
42 26 89.7 509 2 S75721
43 26 89.7 509 2 A12316
44 26 89.7 510 2 S55124
45 26 89.7 515 2 B90504

ALIGNMENTS

RESULT 1

JQ2040
nonstructural protein V - simian paramyxovirus SV41

C;Species: simian paramyxovirus SV41

C;Date: 24-Feb-1994 #sequence_revision 24-Feb-1994 #text_change 16-Jul-1999

C;Accession: JQ2040

R;Kawano, M.; Taurudome, M.; Oki, N.; Nishio, M.; Komada, H.; Matsumura, H.; Kusagawa, J. Gen. Virol. 74, 911-916, 1993

A;Title: Sequence determination of the P gene of simian virus 41: presence of irregular

A;Reference number: JQ2040; MUID:93260408; PMID:8492098

A;Accession: JQ2040

A;Molecule type: genomic RNA

A;Residues: 1-225 <RAW>

A;Cross-references: GB:S60811; NID:G385516; PIDN:AAB26639.1; FID:G385517

C;Genetics:

A;Gene: V

C;Superfamily: simian paramyxovirus nonstructural protein V; V/P protein homology

C;Keywords: alternative splicing; nonstructural protein

F;1-163/Domain: V/P protein homology <VPH>

Query Match 100.0%; Score 29; DB 1; Length 225;

Best Local Similarity 100.0%; Pred. No. 68;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SSGPSL 6

DB 144 SSGPSL 149

RESULT 2

A53027

transcription factor HES-1 - human

N;Alternate names: hairy protein homolog

C;Species: Homo sapiens (man)

C;Date: 03-May-1994 #sequence_revision 03-May-1994 #text_change 03-Dec-1999

C;Accession: A53027

R;Feder, J.N.; Li, L.; Jan, L.Y.; Jan, Y.N.

Genomics 20, 56-61, 1994

A;Title: Genomic cloning and chromosomal localization of HRY, the human homolog to the

A;Reference number: A53027; MUID:94292187; PMID:8020957

A;Accession: A53027

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-280 <FED>

A;Cross-references: GB:L19314; NID:G436999; PIDN:AAA65220.1; FID:G780128

C;Genetics:

A;Gene: GDB:HRV

A;Cross-references: GDB:217077; OMIM:139605

A;Map position: 3q28-3q29

C;Keywords: immediate-early protein

Query Match 100.0%; Score 29; DB 2; Length 280;

```
Best Local Similarity 100.0%; Pred. No. 85;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SSGPSL 6
    |||||
Db 264 SSGPSL 269

RESULT 3
AB3260
Hypothetical membrane spanning protein BMEI0063 [imported] - Brucella melitensis (strain
C;Species: Brucella melitensis
C;Date: 01-Feb-2002 #sequence_revision 01-Feb-2002 #text_change 01-Feb-2002
C;Accession: AB3260
R;DelVecchio, V.G.; Kapratl, V.; Redkar, R.J.; Patra, G.; Mujer, C.; Los, T.; Ivanova,
.; Mazur, M.; Goltzman, E.; Selkov, E.; Eizer, P.H.; Hagius, S.; O'Callaghan, D.; Letess
Proc. Natl. Acad. Sci. U.S.A. 99, 443-448, 2002
A;Title: The genome sequence of the facultative intracellular pathogen Brucella melitens
A;Reference number: AD3252; PMID:11756688
A;Accession: AB3260
A;Status: Preliminary
A;Molecule type: DNA
A;Residues: 1-348 <KUR>
A;Cross-references: GB:AE008917; PIDN:AAL51245.1; PID:gl7981935; GSPDB:GN00190
A;Experimental source: strain 16M
C;Genetics:
A;Gene: BMEI0063
A;Map position: 1

Query Match 100.0%; Score 29; DB 2; Length 348;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SSGPSL 6
    |||||
Db 140 SSGPSL 145

RESULT 4
JQ2041
polymerase-associated nucleocapsid phosphoprotein - simian paramyxovirus SV41
N;Alternate names: P protein
C;Species: simian paramyxovirus SV41
C;Date: 03-Feb-1994 #sequence_revision 03-Feb-1994 #text_change 16-Jul-1999
R;Kawano, M.; Tsurudome, M.; Oki, N.; Nishio, M.; Komada, H.; Matsumura, H.; Kusagawa, S
J. Gen. Virol. 74, 911-916, 1993
A;Title: Sequence determination of the P gene of simian virus 41: presence of irregular
A;Reference number: JQ2040; MUID:93260408; PMID:8492098
A;Accession: JQ2041
A;Molecule type: mRNA
A;Residues: 1-394 <KAW>
C;Genetics:
A;Gene: P
C;Superfamily: simian paramyxovirus P protein; V/P protein homology
C;Keywords: nucleocapsid; phosphoprotein; RNA editing
F;1-163/Domain: V/P protein homology <VPN>

Query Match 100.0%; Score 29; DB 2; Length 394;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SSGPSL 6
    |||||
Db 144 SSGPSL 149

RESULT 5
B97171
uncharacterized conserved protein CAC2200 [imported] - Clostridium acetobutylicum
C;Species: Clostridium acetobutylicum
C;Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 14-Sep-2001
C;Accession: B97171

R;Nolling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee
.; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.
J. Bacteriol. 183, 4823-4838, 2001
A;Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Cl.
A;Reference number: A96900; MUID:21359325; PMID:21359325
A;Accession: B97171
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-405 <KUR>
A;Cross-references: GB:AE001437; PIDN:AAK80157.1; PID:gl5025196; GSPDB:GN00168
A;Experimental source: Clostridium acetobutylicum ATCC824
C;Genetics:
A;Gene: CAC2200

Query Match 100.0%; Score 29; DB 2; Length 405;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SSGPSL 6
    |||||
Db 198 SSGPSL 203

RESULT 6
S74453
Hypothetical protein slr1484 - Synechocystis sp. (strain PCC 6803)
C;Species: Synechocystis sp.
A;Variety: PCC 6803
C;Date: 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 08-Oct-1999
R;Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.
O. K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasud
DNA Res. 3, 109-136, 1996
A;Title: Sequence analysis of the genome of the unicellular cyanobacterium Synechocysti
s.
A;Reference number: S74322; MUID:97061201; PMID:8905231
A;Accession: S74453
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-532 <KAN>
A;Cross-references: EMBL:D90899; GB:AB001339; NID:gl651650; PIDN:BAA16605.1; PID:d10173
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996

Query Match 100.0%; Score 29; DB 2; Length 532;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SSGPSL 6
    |||||
Db 276 SSGPSL 281

RESULT 7
T46659
period clock protein frq [imported] - Neurospora crassa
N;Alternate names: central clock component frq
C;Species: Neurospora crassa
C;Date: 18-Feb-2000 #sequence_revision 18-Feb-2000 #text_change 01-Dec-2000
C;Accession: T46659; S04653
R;Aronson, B.D.; Johnson, K.A.; Dunlap, J.C.
Proc. Natl. Acad. Sci. U.S.A. 91, 7683-7687, 1994
A;Title: Circadian clock locus frequency: protein encoded by a single open reading fram
A;Reference number: Z23120; MUID:94323580; PMID:8052643
A;Accession: T46659
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: mRNA
A;Residues: 1-989 <ARO>
A;Cross-references: EMBL:U17073; NID:9595951; PIDN:AAA57121.1; PID:9595952
A;Experimental source: strain OR 74 A; isolate bda; mycelia
R;McClung, C.R.; Fox, B.A.; Dunlap, J.C.
Nature 339, 558-562, 1989
A;Title: The Neurospora clock gene frequency shares a sequence element with the Drosoph
A;Reference number: S04653; MUID:89281721; PMID:2525233
```

A/Accession: S04653
 A/Status: not compared with conceptual translation
 A/Molecule type: DNA
 A/Residues: 'MCGN', 206, 'P', 208-232, 'H', 234-989 <MCC>
 A/Note: the authors translated the initiation codon GTG as Val
 C/Genetics:
 A/Genes: frq
 A/Map position: VII R

Query Match 100.0%; Score 29; DB 2; Length 989;
 Best Local Similarity 100.0%; Pred. No. 2.9e+02;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SSGPSL 6
 |||||
 Db 271 SSGPSL 276

RESULT 8

S4457
 C/Species: Sordaria fimicola
 C/Date: 28-Oct-1994 #sequence_revision 28-Oct-1994 #text_change 05-Nov-1999
 R/Merrow, M.W.; Dunlap, J.C.
 A/Accession: S4457
 A/Title: Intergenic complementation of a circadian rhythmicity defect: phylogenetic co
 A/Reference number: S4457; PMID:8194516
 A/Accession: S4457
 A/Molecule type: DNA
 A/Residues: 1-997 <MER>
 A/Cross-references: EMBL:L14467; NID:G310366; PIDN:AAA20825.1; PID:G530050
 C/Genetics:
 A/Genes: frq
 C/Keywords: circadian rhythm

Query Match 100.0%; Score 29; DB 2; Length 997;
 Best Local Similarity 100.0%; Pred. No. 2.9e+02;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SSGPSL 6
 |||||
 Db 277 SSGPSL 282

RESULT 9

S30385
 C/Species: Homo sapiens (man)
 C/Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 21-Jan-2000
 R/Milner, C.M.; Campbell, R.D.
 Biochem. J. 290, 811-818, 1993
 A/Title: The G9a gene in the human major histocompatibility complex encodes a novel prob
 A/Reference number: S30385; PMID:8457211
 A/Accession: S30385
 A/Status: preliminary
 A/Molecule type: mRNA
 A/Residues: 1-1001 <MIL>
 A/Cross-references: EMBL:X69838; NID:G287864; PIDN:CAA49491.1; PID:G287865
 C/Superfamily: unassigned ankyrin repeat proteins; ankyrin repeat homology; EGF homology
 F:641-673/Domain: ankyrin repeat homology <ANR>

Query Match 100.0%; Score 29; DB 2; Length 1001;
 Best Local Similarity 100.0%; Pred. No. 2.9e+02;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SSGPSL 6
 |||||
 Db 396 SSGPSL 401

RESULT 10

T04777
 C/Species: Arabidopsis thaliana (mouse-ear cress)
 C/Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 24-Nov-1999
 R/Bevan, M.; De Haan, M.; Maarse, A.C.; Grivell, L.A.; Hobeisel, J.; Mewes, H.W.; Maye
 submitted to the Protein Sequence Database, February 1999
 A/Reference number: Z15384
 A/Accession: T04777
 A/Molecule type: DNA
 A/Residues: 1-1074 <BSV>
 A/Cross-references: EMBL:AL035521
 A/Experimental source: cultivar Columbia; BAC clone F10M10
 C/Genetics:
 A/Map position: 4
 A/Introns: 253/3; 636/3; 1017/3
 A/Note: F10M10.80
 C/Superfamily: Arabidopsis thaliana hypothetical protein F10M10.80

Query Match 100.0%; Score 29; DB 2; Length 1074;
 Best Local Similarity 100.0%; Pred. No. 3.2e+02;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SSGPSL 6
 |||||
 Db 34 SSGPSL 39

RESULT 11

S57150
 C/Species: Saccharomyces cerevisiae
 C/Date: 03-Sep-1995 #sequence_revision 01-Mar-1996 #text_change 06-Feb-1998
 R/Rose, M.; Koetter, P.; Entian, K.D.
 submitted to the Protein Sequence Database, September 1995
 A/Reference number: S57150
 A/Accession: S57150
 A/Molecule type: DNA
 A/Residues: 1-1380 <ROS>
 A/Cross-references: EMBL:Z49627; NID:G1015856; PID:G1015857; MIPS:YJRI27C
 R/Thomas, D.; Barbey, R.; Surdin-Kerjan, Y.
 submitted to the EMBL Data Library, December 1993
 A/Reference number: S43751
 A/Accession: S43751
 A/Molecule type: DNA
 A/Residues: 'MHTN', 18-1116, 'IF', 1119-1130, 'H', 1132-1142, 'S' <THO>
 A/Cross-references: EMBL:U26506; NID:G432497; PID:G432498
 C/Genetics:
 A/Genes: SGD:ZMS1
 A/Cross-references: SGD:S0003888; MIPS:YJRI27C
 A/Map position: 10R

Query Match 100.0%; Score 29; DB 2; Length 1380;
 Best Local Similarity 100.0%; Pred. No. 4e+02;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SSGPSL 6
 |||||
 Db 458 SSGPSL 463

RESULT 12

T34060
 C/Species: Caenorhabditis elegans
 C/Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 04-Mar-2000
 R/Geisel, C.; Kramer, J.; Smith, A.
 submitted to the EMBL Data Library, May 1997
 A/Description: The sequence of C. elegans cosmid F28B3.
 A/Reference number: Z21469

A/Accession: T34060
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: DNA
A/Residues: 1-222 <GEI>
A/Cross-references: EMBL:AF003136; PIDN:AA93634.1; GSPDB:GN00019; CESP:F28B3.2
A/Experimental source: strain Bristol N2; clone F28B3
C/Genetics:
A/Map position: 1
A/Map position: 1
C/Superfamily: Caenorhabditis elegans hypothetical protein F28B3.2

Query Match 93.1%; Score 27; DB 2; Length 222;
Best Local Similarity 83.3%; Pred. No. 1.8e+02; Indels 0; Gaps 0;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 SSGPSL 6
|||
Db 28 SSGPSI 33

RESULT 13
T02850
Hypothetical protein L1439.2 [imported] - Leishmania major (strain Friedlin)
C/Species: Leishmania major
C/Date: 24-Mar-1999 #sequence_revision 24-Mar-1999 #text_change 19-May-2000
C/Accession: F81462; T02850
R/Mylet, P.J.; Audleman, L.; deVos, T.; Hixson, G.; Kiser, P.; Lemley, C.; Magnus, C.;
Proc. Natl. Acad. Sci. U.S.A. 96, 2902-2906, 1999
A/Title: Leishmania major Friedlin chromosome 1 has an unusual distribution of protein-c
A/Reference number: A81455; MUID:99178987; PMID:10077609
A/Accession: F81462
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-1487 <PYL>
A/Cross-references: GB:AE001274; NID:G3264850; PIDN:AAC24673.1; PID:G2266918; GSPDB:GN00
A/Experimental source: strain MHOM/IL/81/Friedlin
C/Genetics:
A/Map position: 1

Query Match 93.1%; Score 27; DB 2; Length 1487;
Best Local Similarity 83.3%; Pred. No. 1.2e+03;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 SSGPSL 6
|||
Db 1478 SSGPSI 1483

RESULT 14
T27121
Hypothetical protein Y53C10A.9 - Caenorhabditis elegans
C/Species: Caenorhabditis elegans
C/Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 04-Mar-2000
C/Accession: T27121
R/White, S.
submitted to the EMBL Data Library, November 1998
A/Reference number: Z20314
A/Accession: T27121
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: DNA
A/Residues: 1-1564 <WIL>
A/Cross-references: EMBL:AL033536; PIDN:CAA22142.1; CESP:Y53C10A.9
A/Experimental source: clone Y53C10A
C/Genetics:
A/Map position: 1

Query Match 93.1%; Score 27; DB 2; Length 1564;
Best Local Similarity 83.3%; Pred. No. 1.2e+03;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 SSGPSL 6
|||
Db 54 SSGPSI 59

RESULT 15
C38230
Inorganic diphosphatase (EC 3.6.1.1) 1 - beet (fragments)
C/Species: Beta vulgaris (beet)
C/Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 27-Oct-2003
C/Accession: C38230; F38230; F38230
R/Sarafian, V.; Kim, Y.; Poole, R.J.; Rea, P.A.
Proc. Natl. Acad. Sci. U.S.A. 89, 1775-1779, 1992
A/Title: Molecular cloning and sequence of cDNA encoding the pyrophosphate-energized va
A/Reference number: A38230; MUID:92179265; PMID:1311852
A/Accession: C38230
A/Molecule type: protein
A/Residues: 1-16 <SAR1>
A/Note: sequence extracted from NCBI backbone (NCBIP:87192)
A/Accession: F38230
A/Molecule type: protein
A/Residues: 17-28 <SAR2>
A/Note: sequence extracted from NCBI backbone (NCBIP:87194)
A/Accession: F38230
A/Molecule type: protein
A/Residues: 29-52 <SAR3>
A/Experimental source: cultivar Detroit Dark
A/Note: sequence extracted from NCBI backbone (NCBIP:87195)
C/Superfamily: H(+)-translocating inorganic pyrophosphatase
C/Keywords: hydrolase

Query Match 89.7%; Score 26; DB 2; Length 52;
Best Local Similarity 83.3%; Pred. No. 72;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 SSGPSL 6
|||
Db 43 TSGPSL 48

Search completed: March 10, 2004, 09:16:44
Job time : 2.92653 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 10, 2004, 09:16:59 ; Search time 3.14397 Seconds
(without alignments)
268.645 Million cell updates/sec

Title: US-09-848-834A-5

Perfect score: 21

Sequence: 1 GPLS 4

Scoring table: BLOSUM62

Gapop 10.0 , Capext 0.5

Searched: 809742 seqs, 21153259 residues

Total number of hits satisfying chosen parameters: 809742

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:

- 1: /cgn2_6/prodata/2/pubpaa/US07_PUBCOMB.pep.*
- 2: /cgn2_6/prodata/2/pubpaa/PCT_NEW_PUB.pep.*
- 3: /cgn2_6/prodata/2/pubpaa/US06_NEW_PUB.pep.*
- 4: /cgn2_6/prodata/2/pubpaa/US06_PUBCOMB.pep.*
- 5: /cgn2_6/prodata/2/pubpaa/US07_NEW_PUB.pep.*
- 6: /cgn2_6/prodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
- 7: /cgn2_6/prodata/2/pubpaa/US08_NEW_PUB.pep.*
- 8: /cgn2_6/prodata/2/pubpaa/US08_PUBCOMB.pep.*
- 9: /cgn2_6/prodata/2/pubpaa/US09_PUBCOMB.pep.*
- 10: /cgn2_6/prodata/2/pubpaa/US09_PUBCOMB.pep.*
- 11: /cgn2_6/prodata/2/pubpaa/US09C_PUBCOMB.pep.*
- 12: /cgn2_6/prodata/2/pubpaa/US09C_NEW_PUB.pep.*
- 13: /cgn2_6/prodata/2/pubpaa/US10A_PUBCOMB.pep.*
- 14: /cgn2_6/prodata/2/pubpaa/US10A_PUBCOMB.pep.*
- 15: /cgn2_6/prodata/2/pubpaa/US10C_PUBCOMB.pep.*
- 16: /cgn2_6/prodata/2/pubpaa/US10C_PUBCOMB.pep.*
- 17: /cgn2_6/prodata/2/pubpaa/US10_NEW_PUB.pep.*
- 18: /cgn2_6/prodata/2/pubpaa/US60_NEW_PUB.pep.*
- 19: /cgn2_6/prodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	21	100.0	4	9	US-09-848-834A-5
2	21	100.0	4	14	US-10-285-976-224
3	21	100.0	6	9	US-09-848-834A-5
4	21	100.0	6	14	US-10-223-711-4
5	21	100.0	8	9	US-09-848-834A-7
6	21	100.0	9	14	US-10-102-622-2
7	21	100.0	10	15	US-10-371-069-209
8	21	100.0	10	15	US-10-371-645-209
9	21	100.0	10	15	US-10-371-260-209
10	21	100.0	10	15	US-10-373-238-49
11	21	100.0	11	15	US-10-371-069-224
12	21	100.0	11	15	US-10-371-645-224
13	21	100.0	11	15	US-10-371-260-224
14	21	100.0	12	10	US-09-854-385-349
15	21	100.0	14	14	US-10-024-123-10

16	21	100.0	14	14	US-10-083-768-39	Sequence 39, Appl
17	21	100.0	14	14	US-10-083-768-216	Sequence 216, Appl
18	21	100.0	15	9	US-09-953-510-67	Sequence 67, Appl
19	21	100.0	15	9	US-09-953-510-68	Sequence 68, Appl
20	21	100.0	15	11	US-09-953-413-67	Sequence 67, Appl
21	21	100.0	15	11	US-09-953-413-68	Sequence 68, Appl
22	21	100.0	15	14	US-10-147-255-67	Sequence 67, Appl
23	21	100.0	15	14	US-10-147-255-68	Sequence 68, Appl
24	21	100.0	18	14	US-10-225-567A-1363	Sequence 1363, Ap
25	21	100.0	20	9	US-09-864-761-33705	Sequence 33705, A
26	21	100.0	20	9	US-09-864-761-36310	Sequence 36310, A
27	21	100.0	20	10	US-09-962-756-845	Sequence 845, App
28	21	100.0	20	15	US-10-253-471-845	Sequence 845, App
29	21	100.0	20	16	US-10-253-493-845	Sequence 845, App
30	21	100.0	21	13	US-10-153-064-74	Sequence 74, Appl
31	21	100.0	21	14	US-10-153-604A-74	Sequence 74, Appl
32	21	100.0	21	14	US-09-813-153-258	Sequence 258, App
33	21	100.0	26	14	US-10-208-304-5	Sequence 5, Appl
34	21	100.0	28	9	US-09-848-834A-11	Sequence 11, Appl
35	21	100.0	29	9	US-09-864-761-46228	Sequence 46228, A
36	21	100.0	30	10	US-09-974-873-334	Sequence 334, App
37	21	100.0	30	10	US-09-563-222-154	Sequence 154, App
38	21	100.0	30	10	US-09-305-736-334	Sequence 334, App
39	21	100.0	30	11	US-09-818-683-334	Sequence 334, App
40	21	100.0	31	9	US-09-848-834A-9	Sequence 9, Appl
41	21	100.0	31	9	US-09-848-834A-15	Sequence 15, Appl
42	21	100.0	31	14	US-10-029-386-30958	Sequence 30958, A
43	21	100.0	33	9	US-09-848-834A-12	Sequence 12, Appl
44	21	100.0	33	14	US-10-012-952A-213	Sequence 213, App
45	21	100.0	34	9	US-09-848-834A-10	Sequence 10, Appl

ALIGNMENTS

RESULT 1
US-09-848-834A-5
; Sequence 5, Application US/09848834A
; Patent No. US20020076416A1
; GENERAL INFORMATION:
; APPLICANT: Aphton Corporation
; TITLE OF INVENTION: Chimeric Peptide Immunogens
; FILE REFERENCE: 1102865-0047
; CURRENT APPLICATION NUMBER: US/09/848,834A
; CURRENT FILING DATE: 2001-05-04
; PRIOR APPLICATION NUMBER: 60/202,328
; PRIOR FILING DATE: 2000-05-05
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 5
; LENGTH: 4
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic peptide
US-09-848-834A-5

Query Match 100.0%; Score 21; DB 9; Length 4;
Best Local Similarity 100.0%; Pred. No. 7.1e+05;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GPLS 4

Db 1 GPLS 4

RESULT 2

US-10-285-976-224
; Sequence 224, Application US/10285976
; Publication No. US20030165500A1
; GENERAL INFORMATION:
; APPLICANT: Rhee, Chae-Seo
; APPLICANT: Malini, Sen


```

; APPLICANT: Wu, Christina
; APPLICANT: Leoni, Lorenzo M.
; APPLICANT: Carr, Maripat
; APPLICANT: Carson, Dennis A.
; APPLICANT: The Regents of the University of California
; TITLE OF INVENTION: Wnt and Frizzled Receptors as Targets for Immunotherapy
; TITLE OF INVENTION: in Head and Neck Squamous Cell Carcinomas
; FILE REFERENCE: 023070-130320US
; CURRENT APPLICATION NUMBER: US/10/285,976
; CURRENT FILING DATE: 2002-11-01
; PRIOR APPLICATION NUMBER: US 60/287,995
; PRIOR FILING DATE: 2001-05-01
; PRIOR APPLICATION NUMBER: WO PCT/US02/13802
; PRIOR FILING DATE: 2002-05-01
; NUMBER OF SEQ ID NOS: 232
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 224
; LENGTH: 4
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: short linker
; OTHER INFORMATION: sequence
US-10-285-976-224

```

```

Query Match      100.0%; Score 21; DB 14; Length 4;
Best Local Similarity 100.0%; Pred. No. 7.1e+05;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy 1 GPSL 4
Db 1 GPSL 4

```

```

RESULT 3
US-09-848-834A-6
; Sequence 6, Application US/09848834A
; Patent No. US20020076416A1
; GENERAL INFORMATION:
; APPLICANT: Apton Corporation
; TITLE OF INVENTION: Chimeric Peptide Immunogens
; FILE REFERENCE: 1102865-0047
; CURRENT APPLICATION NUMBER: US/09/848,834A
; CURRENT FILING DATE: 2001-05-04
; PRIOR APPLICATION NUMBER: 60/202,328
; PRIOR FILING DATE: 2000-05-05
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 6
; LENGTH: 6
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic peptide
US-09-848-834A-6

```

```

Query Match      100.0%; Score 21; DB 9; Length 6;
Best Local Similarity 100.0%; Pred. No. 7.1e+05;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy 1 GPSL 4
Db 3 GPSL 6

```

```

RESULT 4
US-10-223-711-4
; Sequence 4, Application US/10223711
; Publication No. US2003011334A1
; GENERAL INFORMATION:
; APPLICANT: Bakaletz, Lauren O.
; APPLICANT: Kamaya, Pravin T.P.
; TITLE OF INVENTION: Synthetic Chimeric Fimbrin Peptides

```

```

; FILE REFERENCE: 18525/04058
; CURRENT APPLICATION NUMBER: US/10/223,711
; CURRENT FILING DATE: 2002-08-19
; PRIOR APPLICATION NUMBER: 09/148,711
; PRIOR FILING DATE: 1998-09-04
; PRIOR APPLICATION NUMBER: 08/460,502
; PRIOR FILING DATE: 1995-06-02
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 4
; LENGTH: 6
; TYPE: PRT
; ORGANISM: Artificial
; FEATURE:
; OTHER INFORMATION: synthetic construct
US-10-223-711-4

```

```

Query Match      100.0%; Score 21; DB 14; Length 6;
Best Local Similarity 100.0%; Pred. No. 7.1e+05;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy 1 GPSL 4
Db 1 GPSL 4

```

```

RESULT 5
US-09-848-834A-7
; Sequence 7, Application US/09848834A
; Patent No. US20020076416A1
; GENERAL INFORMATION:
; APPLICANT: Apton Corporation
; TITLE OF INVENTION: Chimeric Peptide Immunogens
; FILE REFERENCE: 1102865-0047
; CURRENT APPLICATION NUMBER: US/09/848,834A
; CURRENT FILING DATE: 2001-05-04
; PRIOR APPLICATION NUMBER: 60/202,328
; PRIOR FILING DATE: 2000-05-05
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 7
; LENGTH: 8
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic peptide
US-09-848-834A-7

```

```

Query Match      100.0%; Score 21; DB 9; Length 8;
Best Local Similarity 100.0%; Pred. No. 7.1e+05;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy 1 GPSL 4
Db 3 GPSL 6

```

```

RESULT 6
US-10-102-622-2
; Sequence 2, Application US/10102622
; Publication No. US20030190308A1
; GENERAL INFORMATION:
; APPLICANT: Braun, Ralph P.
; APPLICANT: Thomsen, Lindy
; APPLICANT: Van-Wely, Catherine
; APPLICANT: Ertl, Peter
; TITLE OF INVENTION: Adjuvant
; FILE REFERENCE: 033267-015
; CURRENT APPLICATION NUMBER: US/10/102,622
; CURRENT FILING DATE: 2002-03-19
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2

```

```
; LENGTH: 9
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: HSV CD8 in BALB/C mice
US-10-102-622-2

Query Match      100.0%; Score 21; DB 14; Length 9;
Best Local Similarity 100.0%; Pred. No. 7.1e+05;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 GPSL 4
      ||||
Db      2 GPSL 5

RESULT 7
US-10-371-069-209
; Sequence 209, Application US/10371069
; Publication No. US20030216342A1
; GENERAL INFORMATION:
; APPLICANT: EPIMUNE Inc.
; APPLICANT: Fikes, John D.
; APPLICANT: Hermanson, Gary G.
; APPLICANT: Sette, Alessandro
; APPLICANT: Ishioka, Glenn Y.
; APPLICANT: Livingston, Brian
; APPLICANT: Chesnut, Robert W.
; APPLICANT: Epimmune Inc.
; TITLE OF INVENTION: Expression Vectors for Stimulating an
; TITLE OF INVENTION: Immune Response and Methods of Using the Same
; FILE REFERENCE: 39963-20022.10
; CURRENT APPLICATION NUMBER: US/10/371,069
; CURRENT FILING DATE: 2003-02-21
; PRIOR APPLICATION NUMBER: US 09/078,904
; PRIOR FILING DATE: 1998-05-13
; PRIOR APPLICATION NUMBER: US 60/085,751
; PRIOR FILING DATE: 1998-05-15
; NUMBER OF SEQ ID NOS: 463
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 209
; LENGTH: 10
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: HBV 360 (peptide 1039.01)
US-10-371-069-209

Query Match      100.0%; Score 21; DB 15; Length 10;
Best Local Similarity 100.0%; Pred. No. 3.7e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 GPSL 4
      ||||
Db      6 GPSL 9

RESULT 8
US-10-371-645-209
; Sequence 209, Application US/10371645
; Publication No. US20030216343A1
; GENERAL INFORMATION:
; APPLICANT: EPIMUNE Inc.
; APPLICANT: Fikes, John D.
; APPLICANT: Hermanson, Gary G.
; APPLICANT: Sette, Alessandro
; APPLICANT: Ishioka, Glenn Y.
; APPLICANT: Livingston, Brian
; APPLICANT: Chesnut, Robert W.
; APPLICANT: Epimmune Inc.
; TITLE OF INVENTION: Expression Vectors for Stimulating an
; TITLE OF INVENTION: Immune Response and Methods of Using the Same
; FILE REFERENCE: 39963-20022.11
US-10-371-645-209

Query Match      100.0%; Score 21; DB 15; Length 10;
Best Local Similarity 100.0%; Pred. No. 3.7e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 GPSL 4
      ||||
Db      6 GPSL 9

RESULT 9
US-10-371-260-209
; Sequence 209, Application US/10371260
; Publication No. US20030220285A1
; GENERAL INFORMATION:
; APPLICANT: EPIMUNE Inc.
; APPLICANT: Fikes, John D.
; APPLICANT: Hermanson, Gary G.
; APPLICANT: Sette, Alessandro
; APPLICANT: Ishioka, Glenn Y.
; APPLICANT: Livingston, Brian
; APPLICANT: Chesnut, Robert W.
; APPLICANT: Epimmune Inc.
; TITLE OF INVENTION: Expression Vectors for Stimulating an
; TITLE OF INVENTION: Immune Response and Methods of Using the Same
; FILE REFERENCE: 39963-20022.13
; CURRENT APPLICATION NUMBER: US/10/371,260
; CURRENT FILING DATE: 2003-02-21
; PRIOR APPLICATION NUMBER: US 09/078,904
; PRIOR FILING DATE: 1998-05-13
; PRIOR APPLICATION NUMBER: US 60/085,751
; PRIOR FILING DATE: 1998-05-15
; NUMBER OF SEQ ID NOS: 463
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 209
; LENGTH: 10
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: HBV 360 (peptide 1039.01)
US-10-371-260-209

Query Match      100.0%; Score 21; DB 15; Length 10;
Best Local Similarity 100.0%; Pred. No. 3.7e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 GPSL 4
      ||||
Db      6 GPSL 9

RESULT 10
US-10-373-238-49
; Sequence 49, Application US/10373238
; Publication No. US20040014083A1
; GENERAL INFORMATION:
; APPLICANT: General Atomics
; TITLE OF INVENTION: DETECTION OF HETERODUPLEX POLYNUCLEOTIDES USING MUTANT
```

; TITLE OF INVENTION: NUCLEIC ACID REPAIR ENZYMES WITH ATTENUATED CATALYTIC ACTIVITY
 ; FILE REFERENCE: 45699-20004.20
 ; CURRENT APPLICATION NUMBER: US/10/373,238
 ; CURRENT FILING DATE: 2003-02-24
 ; PRIOR APPLICATION NUMBER: 09/S14,016
 ; PRIOR FILING DATE: 2000-02-25
 ; NUMBER OF SEQ ID NOS: 84
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 49
 ; LENGTH: 10
 ; TYPE: PRT
 ; ORGANISM: Artificial Sequence
 ; FEATURE:
 ; OTHER INFORMATION: Description of Artificial Sequence: Pb (III) ion
 ; PUBLICATION INFORMATION:
 ; PATENT DOCUMENT NUMBER: 5,679,548
 ; PATENT FILING DATE: 1993-06-14
 ; PUBLICATION DATE: 1997-10-21
 US-10-373-238-49

Query Match 100.0%; Score 21; DB 15; Length 10;
 Best Local Similarity 100.0%; Pred. No. 3.7e+02;
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GPSSL 4
 Db 2 GPSSL 5

RESULT 11
 US-10-371-069-224
 ; Sequence 224, Application US/10371069
 ; Publication No. US20030216342A1
 ; GENERAL INFORMATION:
 ; APPLICANT: EPIMUNE Inc.
 ; APPLICANT: Fikes, John D.
 ; APPLICANT: Hermanson, Gary G.
 ; APPLICANT: Sette, Alessandro
 ; APPLICANT: Ishioka, Glenn Y.
 ; APPLICANT: Livingston, Brian
 ; APPLICANT: Chesnut, Robert W.
 ; APPLICANT: Epimmune Inc.
 ; TITLE OF INVENTION: Expression Vectors for Stimulating an
 ; FILE REFERENCE: 39963-20022.10
 ; CURRENT APPLICATION NUMBER: US/10/371,069
 ; CURRENT FILING DATE: 2003-02-21
 ; PRIOR FILING DATE: 1998-05-13
 ; PRIOR APPLICATION NUMBER: US 09/078,904
 ; PRIOR FILING DATE: 1998-05-13
 ; PRIOR APPLICATION NUMBER: US 60/085,751
 ; PRIOR FILING DATE: 1998-05-15
 ; NUMBER OF SEQ ID NOS: 463
 ; SOFTWARE: FastSeq for Windows Version 3.0
 ; SEQ ID NO 224
 ; LENGTH: 11
 ; TYPE: PRT
 ; ORGANISM: Artificial Sequence
 ; FEATURE:
 ; OTHER INFORMATION: HBV env 359 (peptide 1039.06)
 US-10-371-069-224

Query Match 100.0%; Score 21; DB 15; Length 11;
 Best Local Similarity 100.0%; Pred. No. 4e+02;
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GPSSL 4
 Db 7 GPSSL 10

RESULT 12
 US-10-371-645-224

; Sequence 224, Application US/10371645
 ; Publication No. US20030216343A1
 ; GENERAL INFORMATION:
 ; APPLICANT: EPIMUNE Inc.
 ; APPLICANT: Fikes, John D.
 ; APPLICANT: Hermanson, Gary G.
 ; APPLICANT: Sette, Alessandro
 ; APPLICANT: Ishioka, Glenn Y.
 ; APPLICANT: Livingston, Brian
 ; APPLICANT: Chesnut, Robert W.
 ; APPLICANT: Epimmune Inc.
 ; TITLE OF INVENTION: Expression Vectors for Stimulating an
 ; FILE REFERENCE: 39963-20022.11
 ; CURRENT APPLICATION NUMBER: US/10/371,645
 ; CURRENT FILING DATE: 2003-06-20
 ; PRIOR FILING DATE: 1998-05-13
 ; PRIOR APPLICATION NUMBER: US 09/078,904
 ; PRIOR FILING DATE: 1998-05-13
 ; PRIOR APPLICATION NUMBER: US 60/085,751
 ; PRIOR FILING DATE: 1998-05-15
 ; NUMBER OF SEQ ID NOS: 463
 ; SOFTWARE: FastSeq for Windows Version 3.0
 ; SEQ ID NO 224
 ; LENGTH: 11
 ; TYPE: PRT
 ; ORGANISM: Artificial Sequence
 ; FEATURE:
 ; OTHER INFORMATION: HBV env 359 (peptide 1039.06)
 US-10-371-645-224

Query Match 100.0%; Score 21; DB 15; Length 11;
 Best Local Similarity 100.0%; Pred. No. 4e+02;
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GPSSL 4
 Db 7 GPSSL 10

RESULT 13
 US-10-371-260-224
 ; Sequence 224, Application US/10371260
 ; Publication No. US2003020285A1
 ; GENERAL INFORMATION:
 ; APPLICANT: EPIMUNE Inc.
 ; APPLICANT: Fikes, John D.
 ; APPLICANT: Hermanson, Gary G.
 ; APPLICANT: Sette, Alessandro
 ; APPLICANT: Ishioka, Glenn Y.
 ; APPLICANT: Livingston, Brian
 ; APPLICANT: Chesnut, Robert W.
 ; APPLICANT: Epimmune Inc.
 ; TITLE OF INVENTION: Expression Vectors for Stimulating an
 ; FILE REFERENCE: 39963-20022.13
 ; CURRENT APPLICATION NUMBER: US/10/371,260
 ; CURRENT FILING DATE: 2003-02-21
 ; PRIOR FILING DATE: 1998-05-13
 ; PRIOR APPLICATION NUMBER: US 09/078,904
 ; PRIOR FILING DATE: 1998-05-13
 ; PRIOR APPLICATION NUMBER: US 60/085,751
 ; PRIOR FILING DATE: 1998-05-15
 ; NUMBER OF SEQ ID NOS: 463
 ; SOFTWARE: FastSeq for Windows Version 3.0
 ; SEQ ID NO 224
 ; LENGTH: 11
 ; TYPE: PRT
 ; ORGANISM: Artificial Sequence
 ; FEATURE:
 ; OTHER INFORMATION: HBV env 359 (peptide 1039.06)
 US-10-371-260-224

Query Match 100.0%; Score 21; DB 15; Length 11;
 Best Local Similarity 100.0%; Pred. No. 4e+02;

Wed Mar 10 10:34:35 2004

us-09-848-834a-5.open.rapb

Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GPSL 4
 Db 7 GPSL 10

Search completed: March 10, 2004, 10:25:47
 Job time : 3.14397 secs

RESULT 14

US-09-954-385-349
 ; Sequence 349, Application US/09954385
 ; Publication No. US20030100467A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Aehle, Wolfgang
 ; APPLICANT: Baldwin, Toby L.
 ; APPLICANT: Van Gastel, Franciscus J.C.
 ; APPLICANT: Janssen, Giselle G.
 ; APPLICANT: Murray, Christopher J.
 ; APPLICANT: Wang, Huaming
 ; APPLICANT: Winetzk, Deborah S.
 ; TITLE OF INVENTION: Binding Phenol Oxidizing Enzyme-peptide
 ; TITLE OF INVENTION: Complexes
 ; FILE REFERENCE: GC690
 ; CURRENT APPLICATION NUMBER: US/09/954,385
 ; CURRENT FILING DATE: 2001-09-12
 ; NUMBER OF SEQ ID NOS: 433
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 349
 ; LENGTH: 12
 ; TYPE: PRT
 ; ORGANISM: Artificial Sequence
 ; FEATURE:
 ; OTHER INFORMATION: binding peptide
 US-09-954-385-349

Query Match 100.0%; Score 21; DB 10; Length 12;
 Best Local Similarity 100.0%; Pred. No. 4.3e+02;
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GPSL 4
 Db 8 GPSL 11

RESULT 15

US-10-024-123-10
 ; Sequence 10, Application US/10024123
 ; Publication No. US20030022263A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Kastan, Michael
 ; APPLICANT: Canman, Christine
 ; APPLICANT: Kim, Seong-Tae
 ; APPLICANT: Lim, Dae-Sik
 ; APPLICANT: St. Jude Children's Research Hospital
 ; TITLE OF INVENTION: ATM Kinase Modulation for Screening and
 ; TITLE OF INVENTION: Therapies
 ; FILE REFERENCE: 2427/1P142
 ; CURRENT APPLICATION NUMBER: US/10/024,123
 ; CURRENT FILING DATE: 2001-12-17
 ; PRIOR APPLICATION NUMBER: 09/400,653
 ; PRIOR FILING DATE: 1999-09-21
 ; PRIOR APPLICATION NUMBER: 09/248,061
 ; PRIOR FILING DATE: 1999-02-10
 ; NUMBER OF SEQ ID NOS: 49
 ; SOFTWARE: FastSeq for Windows Version 3.0
 ; SEQ ID NO 10
 ; LENGTH: 14
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 US-10-024-123-10

Query Match 100.0%; Score 21; DB 14; Length 14;
 Best Local Similarity 100.0%; Pred. No. 5e+02;
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 10, 2004, 08:58:53 ; Search time 0.731517 Seconds
(without alignments)
284.724 Million cell updates/sec

Title: US-09-848-834A-5

Perfect score: 21

Sequence: 1 GPSL 4

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_42:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	21	100.0	46	1	DIUH_DIPPU
2	21	100.0	46	1	DIUH_LOCOMI
3	21	100.0	46	1	DIUH_PERAM
4	21	100.0	46	1	DIUH_PERAM
5	21	100.0	59	1	NU5M_ARTSA
6	21	100.0	79	1	MAUD_MRTFL
7	21	100.0	83	1	NI9M_BOVIN
8	21	100.0	84	1	NI9M_HUMAN
9	21	100.0	84	1	NI9M_MOUSE
10	21	100.0	96	1	CY22_RHOFL
11	21	100.0	96	1	PRGB_HUMAN
12	21	100.0	97	1	RS6_MYCLE
13	21	100.0	99	1	CY22_RHOMO
14	21	100.0	100	1	CY21_RHOFU
15	21	100.0	102	1	CY21_RHOMO
16	21	100.0	105	1	CYC_EUGR
17	21	100.0	106	1	C55I_SYNY3
18	21	100.0	107	1	CY2_RHOGL
19	21	100.0	107	1	CY2_RHOAC
20	21	100.0	108	1	CYC2_ASCSU
21	21	100.0	110	1	NOLE_RHILP
22	21	100.0	113	1	HV47_MOUSE
23	21	100.0	114	1	TLCA_HUMAN
24	21	100.0	117	1	RL22_SYNP6
25	21	100.0	120	1	C550_BACSU
26	21	100.0	121	1	RL22_SYNY3
27	21	100.0	122	1	Y764_XANAC
28	21	100.0	125	1	KCL1A_PIG
29	21	100.0	125	1	NEUI_HUMAN
30	21	100.0	125	1	NEUI_PIG
31	21	100.0	134	1	EPPI_MOUSE
32	21	100.0	138	1	BGAL_AGRD
33	21	100.0	139	1	A85A_MYCMR
34	21	100.0	140	1	CENX_HUMAN
35	21	100.0	141	1	TM10_HUMAN
36	21	100.0	142	1	TM10_PIG
37	21	100.0	142	1	VE72_SCHPO
38	21	100.0	143	1	PER_DROPI
39	21	100.0	151	1	Y06L_BPTA
40	21	100.0	152	1	Y06L_BPTA
41	21	100.0	162	1	COAD_XILPA
42	21	100.0	162	1	COAD_XILFT
43	21	100.0	162	1	RL10_VIBVU
44	21	100.0	164	1	28S3_ARATH
45	21	100.0	164	1	SSB_MYCTU

ALIGNMENTS

RESULT 1
ID DIUH_DIPPU STANDARD; PRT; 46 AA.
AC P82373;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Diuretic hormone class I (Diuretic peptide) (DP) (DH(46)).
OS Diploptera punctata (Pacific beetle cockroach).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Orthopteroidea; Dictyoptera; Blattaria; Blaberoidea;
OC Blaberidae; Diploptera.
OX NCBI_TaxID=6984;
RN [1]
RP SEQUENCE, FUNCTION, AND MASS SPECTROMETRY.
RT Tissue=Brain, and Corpora cardiaca;
RX MEDLINE=20300924; PubMed=10841553;
RA Furuya K., Milchak R.J., Schegg K.M., Zhang J., Tobe S.S., Coast G.M.,
RA Schooley D.A.;
RT "Cockroach diuretic hormones: characterization of a calcitonin-like
peptide in insects.";
RL Proc Natl Acad Sci U.S.A. 97:6469-6474(2000).
CC -!- FUNCTION: Regulation of fluid secretion. Stimulates primary urine
secretion by Malpighian tubules and causes a dose-dependent
stimulation of cAMP levels in the tubules. Has a greater effect on
the transport of Na(+) than K(+) ions. In vitro, has
synergistic effects with the smaller diuretic hormone DH(31) which
co-occurs with it.
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- MASS SPECTROMETRY: MW=5322; MW ERR=0.1; METHOD=Electrospray.
CC -!- SIMILARITY: Belongs to the sauvagine/corticotropin-releasing
factor/urotensin I family.
DR InterPro; IPR000187; corticotropin.
DR InterPro; IPR003621; Diurtc_hormn.
DR Pfam; PF00473; CRF; 1.
DR ProDom; PD014750; Diurtc_hormn; 1.
DR SMART; SM00039; CRF; 1.
DR PROSITE; PS00511; CRF; 1.
KW Hormone; Amidation.
FT MOD_RES 46 46 AMIDATION.
SQ SEQUENCE 46 AA; 5322 MW; 693CD9A16E47F67E CRC64;
Query Match 100.0%; Score 21; DB 1; Length 46;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GPSL 4
Db 4 GPSL 7
RESULT 2
ID DIUH_LOCOMI STANDARD; PRT; 46 AA.
AC P23455;
DT 01-NOV-1991 (Rel. 20, Created)

DI 01-NOV-1991 (Rel. 20, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Diuretic hormone (DH) (Diuretic peptide) (DP).
OS Locusta migratoria (Migratory locust).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Orthopteroidea; Orthoptera; Caelifera; Acridoidea;
OC Acridoidea; Acrididae; Oedipodinae; Locusta.
OX NCBI_TaxID=7004;
RN [1]
RP SEQUENCE
RC TISSUE=Brain;
RX MEDLINE=91378968; PubMed=1654896;
RA Lehmborg E., Ota R.B., Furiya K., King D.S., Applebaum S.W.,
RA Ferenz H.-J., Schooly D.A.;
RT "Identification of a diuretic hormone of Locusta migratoria";
RL Biochem. Biophys. Res. Commun. 179:1036-1041(1991).
RN [2]
RP SEQUENCE
RC TISSUE=Malpighian tubules;
RX MEDLINE=92126231; PubMed=1663363;
RA Kay I., Wheeler C.H., Coast G.M., Totty N.F., Cusinato O., Patel M.,
RA Goldsworthy G.J.;
RT "Characterization of a diuretic peptide from Locusta migratoria";
RL Biol. Chem. Hoppe-Seyler 372:929-934(1991).
CC -!- FUNCTION: Regulation of fluid secretion. Stimulates primary urine
secretion by Malpighian tubules and causes a dose-dependent
stimulation of CAMP levels in the tubules.
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- SIMILARITY: Belongs to the sauvagine/corticotropin-releasing
factor/urotensin I family.
CC PIR: JH0243; SWLQDA.
DR InterPro: IPR000187; corticoliberin.
DR InterPro: IPR003821; Diurtc_hormn.
DR Pfam: PF00473; CRF; 1.
DR ProDom: PD014750; Diurtc_hormn; 1.
DR SMART: SM00039; CRF; 1.
DR PROSITE: PS00511; CRF; 1.
DR Hormone; Amidation.
FT MOD_RES 46
SQ SEQUENCE 46 AA; 5364 MW; E063260E74939763 CRC64;
Query Match 100.0%; Score 21; DB 1; Length 46;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GPSL 4
DB 4 GPSL 7
RESULT 3
DIUH PERAM STANDARD; PRT; 46 AA.
AC P41538;
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Diuretic hormone (DH) (Diuretic peptide) (DP).
OS Periplaneta americana (American cockroach).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Orthopteroidea; Dictyoptera; Blattaria; Blattoidea;
OC Blattidae; Periplaneta.
OX NCBI_TaxID=6978;
RN [1]
RP SEQUENCE
RX MEDLINE=93174045; PubMed=1337794;
RA Kay I., Patel M., Coast G.M., Totty N.F., Mallet A.I.,
RA Goldsworthy G.J.;
RT "Isolation, characterization and biological activity of a CRF-related
diuretic peptide from Periplaneta americana L.";
RL Regul. Pept. 42:111-122(1992).
CC -!- FUNCTION: Regulation of fluid secretion. Stimulates primary urine
secretion by Malpighian tubules and causes a dose-dependent

CC stimulation of CAMP levels in the tubules.
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- SIMILARITY: Belongs to the sauvagine/corticotropin-releasing
factor/urotensin I family.
CC PIR: A48542; A48542.
DR InterPro: IPR000187; corticoliberin.
DR InterPro: IPR003821; Diurtc_hormn.
DR Pfam: PF00473; CRF; 1.
DR ProDom: PD014750; Diurtc_hormn; 1.
DR SMART: SM00039; CRF; 1.
DR PROSITE: PS00511; CRF; 1.
DR Hormone; Amidation.
FT MOD_RES 46
SQ SEQUENCE 46 AA; 5285 MW; 6938D47CB5AF2BA5 CRC64;
Query Match 100.0%; Score 21; DB 1; Length 46;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GPSL 4
DB 4 GPSL 7
RESULT 4
NUSM ARTSA STANDARD; PRT; 59 AA.
ID NU5M ARTSA
AC P19047;
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-NOV-1990 (Rel. 16, Last sequence update)
DT 01-FEB-1994 (Rel. 28, Last annotation update)
DE NADH-ubiquinone oxidoreductase chain 5 (ZC 1.6.5.3) (Fragment).
GN ND5.
OS Artemia salina (Brine shrimp).
OC Mitochondrion.
OC Eukaryota; Metazoa; Arthropoda; Crustacea; Branchiopoda; Anostraca;
OC Artemiidae; Artemia.
OX NCBI_TaxID=85549;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88289417; PubMed=3135541;
RA Batucacas B., Garesse R., Calleja M., Valverde J.R., Marco R.;
RT "Genome organization of Artemia mitochondrial DNA";
RL Nucleic Acids Res. 16:6515-6529(1988).
CC -!- CATALYTIC ACTIVITY: NADH + ubiquinone = NAD(+) + ubiquinol.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to licenses@isb-sib.ch).
CC -----
DR EMBL: X07663; CAA30510.1;
KW Oxidoreductase; NAD; Ubiquinone; Mitochondrion.
FT NON_TER 1
SQ SEQUENCE 59 AA; 6585 MW; 60C9B6E04827DC23 CRC64;
Query Match 100.0%; Score 21; DB 1; Length 59;
Best Local Similarity 100.0%; Pred. No. 1.7e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GPSL 4
DB 18 GPSL 21
RESULT 5
MAUD METFL STANDARD; PRT; 79 AA.
ID MAUD METFL
AC Q50416; Q50424; Q9RAN3;
DT 01-NOV-1997 (Rel. 35, Created)

01-NOV-1997 (Rel. 35, Last sequence update)
 28-FEB-2003 (Rel. 41, Last annotation update)
 Methylobacillus flagellatum MAUD.
 Bacteria; Proteobacteria; Betaproteobacteria; Methylobacterales;
 Methylobacteriaceae; Methylobacillus.
 NCBI_TaxID=405;
 [1]
 SEQUENCE FROM N.A.
 RP STRAIN=KT / ATCC 51484 / DSM 6875 / VRM B-1610;
 RC MEDLINE=95362636; PubMed=7635847;
 RA Gak E.R., Chistoserdov A.Y., Lidstrom M.E.;
 RT "Cloning, sequencing, and mutation of a gene for acurin in
 Methylobacillus flagellatum KT";
 RL J. Bacteriol. 177:4575-4578 (1995).
 RN [2]
 RP SEQUENCE OF 50-79 FROM N.A.
 RC STRAIN=KT / ATCC 51484 / DSM 6875 / VRM B-1610;
 RA Zhang X., Chistoserdov A.Y., McIntire W.S.;
 RL Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: MAY BE SPECIFICALLY INVOLVED IN THE PROCESSING,
 TRANSPORT, AND/OR MATURATION OF THE MADH BETA-SUBUNIT.
 CC -1- PATHWAY: Methylobacillus utilization.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; L37436; AAC1482.1; -;
 DR EMBL; L37427; AAC1473.1; -;
 DR EMBL; AF114265; AAF03759.1; -;
 KW Transmembrane.
 FT TRANSMEM 5 25 POTENTIAL.
 FT NON CONS 50 51
 SQ SEQUENCE 79 AA; 8517 MW; B4536D2B4A312CEF CRC64;
 Query Match 100.0%; Score 21; DB 1; Length 79;
 Best Local Similarity 100.0%; Pred. No. 2.2e+02;
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 GPSL 4
 DB 49 GPSL 52
 RESULT 6
 NI9M BOVIN STANDARD; PRT; 83 AA.
 ID NI9M BOVIN
 AC Q02371;
 DT 01-JUL-1993 (Rel. 26, Created)
 DT 01-JUL-1993 (Rel. 26, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE NADH-ubiquinone oxidoreductase B9 subunit (EC 1.6.5.3) (EC 1.6.99.3)
 DE (Complex I-B9) (CI-B9).
 GN NDUF3A.
 OS Bos taurus (Bovine).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
 OC Bovidae; Bovinae; Bos.
 OX NCBI_TaxID=9913;
 RN [1]
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 43-73.
 RC MEDLINE=92389317; PubMed=1518044;
 RA Walker J.E., Arizumi J.M., Dupuis A., Fearnley I.M., Finel M.,
 RA Medd S.M., Pilkington S.J., Runswick M.J., Skehel J.M.;
 RT "Sequences of 20 subunits of NADH-ubiquinone oxidoreductase from
 bovine heart mitochondria. Application of a novel strategy for

sequencing proteins using the polymerase chain reaction.";
 J. Mol. Biol. 226:1051-1072 (1992).
 [2]
 RP ACETYLATION.
 RX MEDLINE=95046811; PubMed=7958365;
 RA Fearnley I.M., Skehel J.M., Walker J.E.;
 RT "Electrospray ionization mass spectrometric analysis of subunits of
 NADH-ubiquinone oxidoreductase (Complex I) from bovine heart
 mitochondria";
 RL Biochem. Soc. Trans. 22:551-555 (1994).
 RN [1]
 CC -1- FUNCTION: Transfer of electrons from NADH to the respiratory
 chain. The immediate electron acceptor for the enzyme is believed
 to be ubiquinone.
 CC -1- CATALYTIC ACTIVITY: NADH + ubiquinone = NAD(+) + ubiquinol.
 CC -1- CATALYTIC ACTIVITY: NADH + acceptor = NAD(+) + reduced acceptor.
 CC -1- SUBUNIT: Mammalian complex I is composed of 45 different subunits.
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Mitochondrial
 inner membrane.
 CC -1- SIMILARITY: BELONGS TO THE COMPLEX I B9 SUBUNIT FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; X63217; CAA44902.1; -;
 DR PIR; S28248; S28248.
 KW Oxidoreductase; Ubiquinone; NAD; Mitochondrion; Transmembrane;
 KW Acetylation.
 FT INIT MET 0 0
 FT MOD RES 1 1 ACETYLATION.
 FT TRANSMEM 17 37 POTENTIAL.
 SQ SEQUENCE 83 AA; 9218 MW; 1B7B031860DB9370 CRC64;
 Query Match 100.0%; Score 21; DB 1; Length 83;
 Best Local Similarity 100.0%; Pred. No. 2.3e+02;
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 GPSL 4
 DB 74 GPSL 77
 RESULT 7
 NI9M HUMAN STANDARD; PRT; 84 AA.
 ID NI9M HUMAN
 AC Q95167;
 DT 15-JUL-1999 (Rel. 38, Created)
 DT 15-JUL-1999 (Rel. 38, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE NADH-ubiquinone oxidoreductase B9 subunit (EC 1.6.5.3) (EC 1.6.99.3)
 DE (Complex I-B9) (CI-B9).
 GN NDUF3A.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=99097250; PubMed=9878551;
 RA Loeffen J.L.C.M., Triepels R.H., van den Heuvel L., Schuelke M.,
 RA Bastens C.A.F., Smeets R.J.P., Trijbels J.M.F., Smeitink J.A.M.;
 RT "cDNA of eight nuclear encoded subunits of NADH:ubiquinone
 oxidoreductase: human complex I cDNA characterization completed.";
 RL Biochem. Biophys. Res. Commun. 253:415-422 (1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Blood;
 RX MEDLINE=20499367; PubMed=11042152;
 Zhang Q.-H., Ye M., Wu X.-Y., Ren S.-X., Zhao M., Zhao C.-J., Fu G.,

Shen Y., Pan H.-Y., Lu G., Zhong M., Xu X.-R., Han Z.-G., Zhang J.-W.,
 Tao J., Huang Q.-H., Zhou J., Hu G.-X., Gu J., Chen S.-J., Chen Z.,
 RT "Cloning and functional analysis of cDNAs with open reading frames for
 RT 300 previously undefined genes expressed in CD34⁺ hematopoietic
 RT stem/progenitor cells";
 RL Genome Res. 10:1546-1560(2000).
 RN [3].
 RP SEQUENCE FROM N.A.
 RC TISSUE=prostate;
 RX MEDLINE=22388257; PubMed=12477932;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Schenck C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan M., Moore T., Max S.I., Wang J., Hsieh F.,
 RA Datchenko L., Nardusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
 RA Rana S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullihy S.J.,
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahey J., Helton E., Kettaman M., Madan A., Rodriguez S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko V., Bouffard G.G.,
 RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
 RA Butterfield Y.S.N., Krzyzanski M.I., Skalska U., Smalish U.A.,
 RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.,
 RA "Generation and initial analysis of more than 15,000 full-length
 RT human and mouse cDNA sequences";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 CC -!- FUNCTION: Transfer of electrons from NADH to the respiratory
 CC chain. The immediate electron acceptor for the enzyme is believed
 CC to be ubiquinone.
 CC -!- CATALYTIC ACTIVITY: NADH + ubiquinone = NAD(+) + ubiquinol.
 CC -!- CATALYTIC ACTIVITY: NADH + acceptor = NAD(+) + reduced acceptor.
 CC -!- SUBUNIT: Mammalian complex I is composed of 45 different subunits.
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein. Mitochondrial
 CC inner membrane.
 CC -!- SIMILARITY: BELONGS TO THE COMPLEX I B9 SUBUNIT FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC at the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 DR EMBL; AF044955; AAD05420.1; -;
 DR EMBL; AF070653; AAD20959.1; -;
 DR EMBL; EC022369; AAB22369.1; -;
 DR F01; J03079; J03079.
 DR Genbank; HGNC:7686; NDUFA3.
 DR MIM; 603832; -;
 DR GO; GO:0008137; F:NADH dehydrogenase (ubiquinone) activity; TAS.
 KW Oxidoreductase; Ubiquinone; NAD; Mitochondrion; Transmembrane.
 FT TRANSMEM 19 39 POTENTIAL.
 SQ SEQUENCE 84 AA; 9279 MW; 38B27A96D7A05D31 CRC64;
 Query Match 100.0%; Score 21; DB 1; Length 84;
 Best Local Similarity 100.0%; Pred. No. 2.3e+02;
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 GPST 4
 Db 75 GPST 78
 RESULT 8
 ID N19M MOUSE STANDARD; PRT; 84 AA.
 AC Q9CQ31;
 DT 28-FEB-2003 (Rel. 41, Created)

28-FEB-2003 (Rel. 41, Last sequence update)
 10-OCT-2003 (Rel. 42, Last annotation update)
 DE NADH-ubiquinone oxidoreductase B9 subunit (EC 1.6.5.3) (EC 1.6.99.3)
 DE (Complex I-B9) (CI-B9).
 GN NDUFA3.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1].
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Heart, and Tongue;
 RX MEDLINE=21085660; PubMed=11217851;
 RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
 RA Azakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
 RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamakawa I.,
 RA Saito T., Okazaki Y., Gojohori T., Bono H., Kasukawa T., Saito R.,
 RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
 RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
 RA Kuehl P., Lewis S., Matsuo Y., Nikola I., Pesole G., Quackenbush J.,
 RA Schirai L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
 RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
 RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
 RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
 RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
 RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
 RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
 RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
 RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Wilming L.,
 RA Wyshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohsaki S.,
 RA Hayashizaki Y.,
 RL "Functional annotation of a full-length mouse cDNA collection";
 RT Nature 409:685-690(2001).
 CC -!- FUNCTION: Transfer of electrons from NADH to the respiratory
 CC chain. The immediate electron acceptor for the enzyme is believed
 CC to be ubiquinone (By similarity).
 CC -!- CATALYTIC ACTIVITY: NADH + ubiquinone = NAD(+) + ubiquinol.
 CC -!- CATALYTIC ACTIVITY: NADH + acceptor = NAD(+) + reduced acceptor.
 CC -!- SUBUNIT: Mammalian complex I is composed of 45 different subunits.
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein. Mitochondrial
 CC inner membrane (By similarity).
 CC -!- SIMILARITY: BELONGS TO THE COMPLEX I B9 SUBUNIT FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC at the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 DR EMBL; AK019099; BAB31545.1; -;
 DR EMBL; AK003133; BAB25993.1; -;
 DR MGD; MGI:1913341; Ndufa3.
 KW Oxidoreductase; Ubiquinone; NAD; Mitochondrion; Transmembrane.
 FT TRANSMEM 19 39 POTENTIAL.
 SQ SEQUENCE 84 AA; 9331 MW; B05E59B217D95A92 CRC64;
 Query Match 100.0%; Score 21; DB 1; Length 84;
 Best Local Similarity 100.0%; Pred. No. 2.3e+02;
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 GPST 4
 Db 75 GPST 78
 RESULT 9
 ID CY22 RHOFU STANDARD; PRT; 96 AA.
 AC P00089;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 21-JUL-1986 (Rel. 01, Last sequence update)


```

DT 01-NOV-1998 (Rel. 09, Last annotation update)
DE Cytochrome C2, iso-2.
OS Rhodospirillum rubrum.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhodospirillales;
OC Rhodospirillaceae; Phaeospirillum.
OX NCBI_TaxID=1082;
RN [1]
RP SEQUENCE.
RX MEDLINE=79199667; PubMed=221822;
RA Ambler R.P., Daniel M., Hermoso J., Meyer T.E., Bartsch R.G.,
RA Kanen M.D.;
RA "Cytochrome c2 sequence variation among the recognised species of
RT purple nonsulphur photosynthetic bacteria.";
RL Nature 278:659-660(1979).
CC -!- FUNCTION: CYTOCHROME C2 IS FOUND MAINLY IN PURPLE, NONSULFUR.
CC PHOTOSYNTHETIC BACTERIA WHERE IT FUNCTIONS AS THE ELECTRON DONOR
CC TO THE OXIDIZED BACTERIOCHLOROPHYLL IN THE PHOTOPHOSPHORYLATION
CC PATHWAY. HOWEVER, IT MAY ALSO HAVE A ROLE IN THE RESPIRATORY CHAIN
CC AND IS FOUND IN SOME NONPHOTOSYNTHETIC BACTERIA.
DR PIR; A00081; CCOFF2.
DR HSSP; P00004; 1WEJ.
DR InterPro; IPR003088; Cyt_C1.
DR InterPro; IPR003327; Cyt_C1AB.
DR InterPro; IPR003345; CytC_heme_BS.
DR Pfam; PF00034; cytochrome c; 1.
DR PRINTS; PR00604; CYTCRMECIAB.
DR ProDom; PD000375; Cyt_C1AB; 1.
DR PROSITE; PS00190; CYTOCHROME C; 1.
DR Electron transport; Photosynthesis; Heme.
FT BINDING 10 10 HEME (COVALENT).
FT BINDING 13 13 HEME (COVALENT).
FT METAL 14 14 IRON (HEME AXIAL LIGAND).
FT METAL 75 75 IRON (HEME AXIAL LIGAND).
FT VARIANT 62 62 Y -> F (IN ABOUT 30% OF THE MOLECULES).
SQ SEQUENCE 96 AA; 9832 MW; E51782ABACBD042D CRC64;

Query Match 100.0%; Score 21; DB 1; Length 96;
Best Local Similarity 100.0%; Pred. No. 2.7e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GPSSL 4 PRT; 96 AA.
Db 24 GPSSL 27

RESULT 10
PRGB HUMAN STANDARD; PRT; 96 AA.
AC Q02325;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Plasmimogen-related protein B precursor.
GN PLGL OR PRGB.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92359990; PubMed=1379800;
RA Weissbach L., Treadwell B.V.;
RT "A plasmimogen-related gene is expressed in cancer cells.";
RL Biochem. Biophys. Res. Commun. 186:1108-1114(1992).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=92207924; PubMed=1554698;
RA Ichinose A.;
RT "Multiple members of the plasmimogen-apolipoprotein(a) gene family
RT associated with thrombosis.";
RL Biochemistry 31:3113-3118(1992).
RN [3]
RP SEQUENCE FROM N.A.

```

```

RC TISSUE=Liver, and Skeletal muscle;
RX MEDLINE=2238257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Heish F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahy J., Hellon E., Kettman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalusz D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length
RT human and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
CC -!- FUNCTION: May bind noncovalently to lysine binding sites
CC present in the kringles structures of plasminogen. This may
CC interfere with the binding of fibrin or alpha-2-antiplasmin
CC to plasminogen and may result in the localization of activity
CC at sites necessary for extracellular matrix destruction.
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- SIMILARITY: Contains 1 PAN domain.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC EMBL; M33143; AAB06491.1; -.
CC EMBL; M86876; AAA60169.1; -.
DR EMBL; M86874; AAA60169.1; JOINED.
DR EMBL; M86875; AAA60169.1; JOINED.
DR EMBL; BC005379; AAH05379.1; -.
DR EMBL; BC022294; AAH22294.1; -.
DR PIR; JCI195; JCI195.
DR Genew; HGNC:9072; PLGL.
DR MIM; 173340; -.
DR InterPro; IPR003014; PAN.
DR InterPro; IPR003603; Pan_app.
DR Pfam; PF00024; PAN; 1.
DR SMART; SM00473; PAN_AP; 1.
KW Signal.
FT SIGNAL 1 19 BY SIMILARITY.
FT CHAIN 20 96 PLASMINOGEN-RELATED PROTEIN B.
FT DISULFID 49 73 BY SIMILARITY.
FT DISULFID 53 61 BY SIMILARITY.
SQ SEQUENCE 96 AA; 10971 MW; 8D8073F529F2716D CRC64;

Query Match 100.0%; Score 21; DB 1; Length 96;
Best Local Similarity 100.0%; Pred. No. 2.7e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GPSSL 4
Db 30 GPSSL 33

RESULT 11
RS6 MYCLE
ID RS6 MYCLE STANDARD; PRT; 96 AA.
AC P46389;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)

```

DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE 30S ribosomal protein S6.
 GN RPSF OR M22685 OR MLCB1913.21C.
 OS Mycobacterium leprae.
 CC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 CC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 OX NCBI_TaxID=1769;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=97124199; PubMed=8969512;
 RA Psihi H., de Rossi E., Salazar L., Cantoni R., Labo M., Riccardi G.,
 RA Takiff H.E., Eigmeier K., Bergh S., Cole S.T.;
 RT "Gene arrangement and organization in an approximately 76 kb fragment
 RT encompassing the oric region of the chromosome of Mycobacterium
 RT leprae";
 RI Microbiology 142:3147-3161 (1996).
 RL [2]
 RN SEQUENCE FROM N.A.
 RP STRAIN=TN;
 RX MEDLINE=21128732; PubMed=11234002;
 RA Cole S.T., Eigmeier K., Parkhill J., James K.D., Thomson N.R.,
 RA Wheeler P.R., Honore D., Garnier T., Churcher C., Harris D.,
 RA Mungall K., Basham D., Brown D., Chillingworth T., Connor K.,
 RA Davies K.M., Devlin K., Duthoy S., Feltwell T., Fraser A., Hamlin N.,
 RA Holtroyd S., Hornaby T., Jagels K., Lacroix C., Maclean J., Moule S.,
 RA Murphy L., Oliver K., Quail M.A., Rajandream M.A., Rutherford K.M.,
 RA Rutter S., Seeger K., Simon S., Simmonds M., Skelton J., Squares R.,
 RA Squares S., Stevens K., Taylor K., Whitehead S., Woodward J.R.,
 RA Barrall B.G.;
 RA "Massive gene decay in the leprosy bacillus";
 RT Nature 409:1007-1011 (2001).
 RL [2]
 RN NATURE 409:1007-1011 (2001).
 RP -!- FUNCTION: Binds together with S18 to 16S ribosomal RNA (By
 CC similarity).
 CC -!- SIMILARITY: Belongs to the SGP family of ribosomal proteins.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; L39923; AAB53121.1; -;
 DR EMBL; AL022118; CAAL7954.1; -;
 DR EMBL; AL583926; CAC32217.1; -;
 DR PIR; C87245; C87245.
 DR Leproma; M22685; -;
 DR HAMAP; MF_00360; -; 1.
 DR InterPro; IPR000529; Ribosomal_S6.
 DR Pfam; PF01250; Ribosomal_S6; 1.
 DR ProDom; PD003809; Ribosomal_S6; 1.
 DR TIGRFAMs; TIGR00166; S6; 1.
 DR PROSITE; PS01048; RIBOSOMAL_S6; 1.
 KW Ribosomal protein; rRNA-binding; Complete proteome.
 SQ SEQUENCE 96 AA; 10868 MW; CAA9C99824B08C2 CRC64;
 Query Match 100.0%; Score 21; DB 1; Length 96;
 Best Local Similarity 100.0%; Pred. No. 2.7e+02;
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 1 GPSSL 4
 DB 20 GPSSL 23
 RESULT 12
 CY22_RHOMO STANDARD; PRT; 97 AA.
 AC P00088;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 21-JUL-1986 (Rel. 01, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)

DE Cytochrome C2, iso-2.
 OS Rhodospirillum rubrum.
 CC Bacteria; Proteobacteria; Alphaproteobacteria; Rhodospirillales;
 CC Rhodospirillaceae; Phaeospirillum.
 OX NCBI_TaxID=1083;
 RN [1]
 RP SEQUENCE.
 RA Ambler R.P., Meyer T.E., Bartsch R.G., Kamen M.D.;
 RA Unpublished results, cited by:
 RL Ambler R.P.;
 RL (In) Matsubara H., Yamanaka T. (eds.);
 RL Evolution of protein molecules, pp.311-322, Japan Scientific Societies
 RL Press/Center for Academic Publications, Tokyo (1978).
 RN [2]
 RP SEQUENCE.
 RA Ambler R.P.;
 RL Submitted (JUN-1977) to the PIR data bank.
 CC -!- FUNCTION: CYTOCHROME C2 IS FOUND MAINLY IN PURPLE, NONSULFUR,
 CC PHOTOSYNTHETIC BACTERIA WHERE IT FUNCTIONS AS THE ELECTRON DONOR
 CC TO THE OXIDIZED BACTERIOCHLOROPHYLL IN THE PHOTOPHOSPHORYLATION
 CC PATHWAY. HOWEVER, IT MAY ALSO HAVE A ROLE IN THE RESPIRATORY CHAIN
 CC AND IS FOUND IN SOME NONPHOTOSYNTHETIC BACTERIA.
 CC PIR; A00080; CCOFMP.
 DR HSSP; P00004; IWEJ.
 DR InterPro; IPR003088; Cyt_C1.
 DR InterPro; IPR002327; Cyt_C1A8.
 DR InterPro; IPR000345; Cyt_C heme BS.
 DR Pfam; PF00034; cytochrome_c_1_
 DR PRINTS; PR00604; CYTCHRMSC1A8.
 DR ProDom; PD000375; Cyt_C1A8; 1.
 DR PROSITE; PS0190; CYTOCHROME_C; 1.
 KW Electron transport; Photosynthesis; Heme.
 FT BINDING 10 10 HEME (COVALENT).
 FT BINDING 13 13 HEME (COVALENT).
 FT METAL 14 14 IRON (HEME AXIAL LIGAND).
 FT METAL 75 75 IRON (HEME AXIAL LIGAND).
 SQ SEQUENCE 97 AA; 10222 MW; 5A1C8C489B0DBF6 CRC64;
 Query Match 100.0%; Score 21; DB 1; Length 97;
 Best Local Similarity 100.0%; Pred. No. 2.7e+02;
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 1 GPSSL 4
 DB 24 GPSSL 27
 RESULT 13
 CY21_RHOFU STANDARD; PRT; 99 AA.
 AC P00086;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 21-JUL-1986 (Rel. 01, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE Cytochrome C2, iso-1.
 OS Rhodospirillum rubrum.
 CC Bacteria; Proteobacteria; Alphaproteobacteria; Rhodospirillales;
 CC Rhodospirillaceae; Phaeospirillum.
 OX NCBI_TaxID=1082;
 RN [1]
 RP SEQUENCE.
 RA Ambler R.P., Meyer T.E., Bartsch R.G., Kamen M.D.;
 RA Unpublished results, cited by:
 RL Ambler R.P.;
 RL (In) Matsubara H., Yamanaka T. (eds.);
 RL Evolution of protein molecules, pp.311-322, Japan Scientific Societies
 RL Press/Center for Academic Publications, Tokyo (1978).
 RN [2]
 RP SEQUENCE.
 RA Ambler R.P.;
 RL Submitted (JUN-1977) to the PIR data bank.
 CC -!- FUNCTION: CYTOCHROME C2 IS FOUND MAINLY IN PURPLE, NONSULFUR,
 CC PHOTOSYNTHETIC BACTERIA WHERE IT FUNCTIONS AS THE ELECTRON DONOR

CC TO THE OXIDIZED BACTERIOCHLOROPHYLL IN THE PHOTOPHOSPHORYLATION
 CC PATHWAY. HOWEVER, IT MAY ALSO HAVE A ROLE IN THE RESPIRATORY CHAIN
 CC AND IS FOUND IN SOME NONPHOTOSYNTHETIC BACTERIA.

CC PIR; A00078; CCOF2F.
 DR HSSP; P00045; 1YEB.
 DR InterPro; IPR003088; Cyt_C1.
 DR InterPro; IPR002327; Cyt_C1AB.
 DR InterPro; IPR000345; CytC_heme_BS.
 DR Pfam; PF00034; cytochrome_c; 1.
 DR PRINTS; PR00604; CYTCHRMCEIAB.
 DR ProDom; PD000375; Cyt_C1AB; 1.
 DR PROSITE; PS00190; CYTOCHROME_C; 1.
 DR Electron transport; Photosynthesis; Heme.
 FT BINDING 10 10 HEME (COVALENT).
 FT BINDING 13 13 HEME (COVALENT).
 FT METAL 14 14 IRON (HEME AXIAL LIGAND).
 FT METAL 75 75 IRON (HEME AXIAL LIGAND).
 SQ SEQUENCE 99 AA; 10288 MW; 10248 MW; 68FE592E3BF19D0 CRC64;

Query Match 100.0%; Score 21; DB 1; Length 99;
 Best Local Similarity 100.0%; Pred. No. 2.8e+02;
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GPSL 4
 DB 24 GPSL 27

RESULT 14
 CY21_RHOMO STANDARD; PRT; 100 AA.
 AC P00087; (Rel. 01, Created)
 DT 21-JUL-1986 (Rel. 01, Last sequence update)
 DT 21-JUL-1986 (Rel. 01, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE Cytochrome C2, iso-1.
 DE Rhodospirillum rubrum.
 OS Bacteria; Proteobacteria; Alphaproteobacteria; Rhodospirillales;
 OC Rhodospirillaceae; Rhodospirillum.
 OX NCBI_TaxID=1093;
 RN [1]
 RP SEQUENCE.
 RA Ambler R.P.; Meyer T.E.; Bartsch R.G.; Kamen M.D.;
 RA Unpublished results, cited by:
 RL Ambler R.P.;
 RL (In) Matsuura H.; Yamanaka T. (eds.);
 RL Evolution of protein molecules, pp 311-322, Japan Scientific Societies
 RL Press/Center for Academic Publications, Tokyo (1978).
 RN [2]
 RP SEQUENCE.

RA Ambler R.P.;
 RA Submitted (JUN-1977) to the PIR data bank.
 CC -1- FUNCTION: CYTOCHROME C2 IS FOUND MAINLY IN PURPLE, NONSULFUR,
 CC PHOTOSYNTHETIC BACTERIA WHERE IT FUNCTIONS AS THE ELECTRON DONOR
 CC TO THE OXIDIZED BACTERIOCHLOROPHYLL IN THE PHOTOPHOSPHORYLATION
 CC PATHWAY. HOWEVER, IT MAY ALSO HAVE A ROLE IN THE RESPIRATORY CHAIN
 CC AND IS FOUND IN SOME NONPHOTOSYNTHETIC BACTERIA.

CC PIR; A00079; CCOF2M.
 DR HSSP; P00044; 1RAC.
 DR InterPro; IPR003088; Cyt_C1.
 DR InterPro; IPR002327; Cyt_C1AB.
 DR InterPro; IPR000345; CytC_heme_BS.
 DR Pfam; PF00034; cytochrome_c; 1.
 DR PRINTS; PR00604; CYTCHRMCEIAB.
 DR ProDom; PD000375; Cyt_C1AB; 1.
 DR PROSITE; PS00190; CYTOCHROME_C; 1.
 DR Electron transport; Photosynthesis; Heme.
 FT BINDING 11 11 HEME (COVALENT).
 FT BINDING 14 14 HEME (COVALENT).
 FT METAL 15 15 IRON (HEME AXIAL LIGAND).
 FT METAL 76 76 IRON (HEME AXIAL LIGAND).
 SQ SEQUENCE 100 AA; 10248 MW; 68FE592E3BF19D0 CRC64;

Query Match 100.0%; Score 21; DB 1; Length 100;
 Best Local Similarity 100.0%; Pred. No. 2.8e+02;
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GPSL 4
 DB 25 GPSL 28

RESULT 15
 CYC_EUGGR STANDARD; PRT; 102 AA.
 ID CYC_EUGGR STANDARD; PRT; 102 AA.
 AC P00076;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 21-JUL-1986 (Rel. 01, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Cytochrome c.
 DE Euglena gracilis.
 OS Eukaryota; Euglenozoa; Euglenida; Euglenales; Euglena.
 OX NCBI_TaxID=3039;
 RN [1]
 RP SEQUENCE.
 RX MEDLINE=76039443; PubMed=170910;
 RA Pettigrew G.W.; Leaver J.L.; Meyer T.E.; Ryle A.P.;
 RT "Purification, properties and amino acid sequence of atypical
 RT cytochrome c from two protozoa, Euglena gracilis and Crithidia
 RT oncopelti";
 RL Biochem. J. 147:291-302 (1975).
 CC -1- FUNCTION: Electron carrier protein. The oxidized form of the
 CC cytochrome c heme group can accept an electron from the heme group
 CC of the cytochrome c1 subunit of cytochrome reductase. Cytochrome c
 CC then transfers this electron to the cytochrome oxidase complex,
 CC the final protein carrier in the mitochondrial electron-transport
 CC chain.
 CC -1- SUBCELLULAR LOCATION: Mitochondrial matrix.
 CC -1- PTM: Binds 1 heme group per molecule.
 CC -1- MISCELLANEOUS: Compared to other cytochrome c's lacks one of the
 CC two cysteines that covalently bind the heme group.
 CC -1- SIMILARITY: Belongs to the cytochrome c family.

CC PIR; A00068; CCEG.
 DR HSSP; P00004; 1MEJ.
 DR InterPro; IPR003088; Cyt_C1.
 DR InterPro; IPR002327; Cyt_C1AB.
 DR InterPro; IPR000345; CytC_heme_BS.
 DR Pfam; PF00034; cytochrome_c; 1.
 DR PRINTS; PR00604; CYTCHRMCEIAB.
 DR ProDom; PD000375; Cyt_C1AB; 1.
 DR PROSITE; PS00190; CYTOCHROME_C; FALSE_NEG.
 DR Mitochondrion; Electron transport; Respiratory chain; Heme;
 KW Methylation; Acetylation.
 FT MOD_RES 1 1 ACETYLATION.
 FT BINDING 17 17 HEME (COVALENT).
 FT METAL 18 18 IRON (HEME AXIAL LIGAND).
 FT METAL 79 79 IRON (HEME AXIAL LIGAND).
 FT MOD_RES 85 85 METHYLATION (TRI-).
 SQ SEQUENCE 102 AA; 11210 MW; 78414AD11BDA6F84 CRC64;

Query Match 100.0%; Score 21; DB 1; Length 102;
 Best Local Similarity 100.0%; Pred. No. 2.8e+02;
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GPSL 4
 DB 28 GPSL 31

Search completed: March 10, 2004, 09:13:50
 Job time : 1.73152 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 10, 2004, 08:58:54 ; Search time 1.24514 Seconds
(without alignments)
309.015 Million cell updates/sec

Title: US-09-848-834A-5

Perfect score: 21

Sequence: 1 GPSL 4

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 293366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 293366

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR 78:*

1: Pir1.*

2: Pir2.*

3: Pir3.*

4: Pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	21	100.0	14	2 A43847	IG mu chain V regi
2	21	100.0	27	2 P00583	tyrosine 3-monooxy
3	21	100.0	27	2 P00584	tyrosine 3-monooxy
4	21	100.0	27	2 P00585	tyrosine 3-monooxy
5	21	100.0	27	2 P00586	tyrosine 3-monooxy
6	21	100.0	32	2 S20771	IG heavy chain V r
7	21	100.0	35	2 C25159	hypothetical prote
8	21	100.0	46	1 SWLQDA	diuretic hormone -
9	21	100.0	46	2 A48542	hypothetical prote
10	21	100.0	50	2 P30584	hypothetical prote
11	21	100.0	52	2 C38230	incrganic diphosph
12	21	100.0	54	2 S35628	myosin heavy chain
13	21	100.0	54	2 T11181	hypothetical prote
14	21	100.0	56	2 T03658	phosphoenolpyruvat
15	21	100.0	56	2 S36380	IG heavy chain V r
16	21	100.0	58	2 D70502	hypothetical prote
17	21	100.0	58	2 S34413	NMA protein - hum
18	21	100.0	59	2 S01877	NADH2 dehydrogenas
19	21	100.0	66	2 T36517	hypothetical prote
20	21	100.0	68	2 T36853	hypothetical prote
21	21	100.0	73	2 T30612	hypothetical prote
22	21	100.0	75	2 B24853	hypothetical prote
23	21	100.0	80	2 B36780	hypothetical prote
24	21	100.0	80	2 D36861	hypothetical prote
25	21	100.0	82	2 C35296	transcription regu
26	21	100.0	84	2 J50379	NADH2 dehydrogenas
27	21	100.0	84	2 B28248	NADH2 dehydrogenas
28	21	100.0	85	2 B69770	conserved hypothet
29	21	100.0	91	2 JQ2204	hypothetical 10.7K

30	21	100.0	96	1 CCQPF2	cytochrome c2, iso
31	21	100.0	96	1 JCL195	plasminogen-relate
32	21	100.0	96	2 C87245	30S ribosomal prot
33	21	100.0	96	2 PC4212	gastrolith matrix
34	21	100.0	97	1 CCQFM2	cytochrome c2, iso
35	21	100.0	97	1 D89861	Na+/H+ antiporter
36	21	100.0	99	1 CCQPF2	cytochrome c2, iso
37	21	100.0	100	1 CCQFM2	cytochrome c2, iso
38	21	100.0	100	2 S14485	IG heavy chain V r
39	21	100.0	101	2 S14484	IG heavy chain V r
40	21	100.0	101	2 B87208	probable secreted
41	21	100.0	102	1 CCEG	cytochrome c [vali
42	21	100.0	102	2 S14486	IG heavy chain V r
43	21	100.0	102	2 S14487	IG heavy chain V r
44	21	100.0	102	2 S14488	IG heavy chain V r
45	21	100.0	102	2 H72609	hypothetical prote

ALIGNMENTS

RESULT 1

A43847
IG mu chain V region - rabbit (fragment)
C:Species: Oryctolagus cuniculus (domestic rabbit)
C:Date: 17-Apr-1993 #sequence_revision 30-Sep-1993 #text_change 16-Aug-1996
C:Accession: A43847; A37272
R:Porward, D.W.; Huguenel, E.D.; Davis, G.; Garon, C.F.
A:Title: Interactions between extracellular Borrelia burgdorferi proteins and non-Borr
Infect. Immun. 60, 838-844, 1992
A:Reference number: A43847; MUID:92175997; PMID:1541558
A:Accession: A43847
A:Molecule type: protein
A:Residues: 7, 2, 14 <DOR>
A:Note: sequence extracted from NCBI backbone (NCBI:85360)
A:Note: The identifications of the protein and of the source were confirmed by the aut
C:Keywords: immunoglobulin; pyroglutamic acid
F:1/Modified site: pyroglutamate carboxylic acid (Gln) #status experimental

Query Match 100.0%; Score 21; DB 2; Length 14;
Best Local Similarity 100.0%; Pred. No. 88;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GPSL 4

DB 8 GPSL 11

RESULT 2

PN0583
tyrosine 3-monooxygenase (EC 1.14.16.2) type 3 and 4 - chimpanzee (fragment)
N:Alternate names: oxygen oxidoreductase; tetrahydropteridine; tyrosine hydroxylase
C:Species: Pan troglodytes (chimpanzee)
C:Date: 03-May-1994 #sequence_revision 07-Oct-1994 #text_change 31-Mar-2000
C:Accession: PN0583
R:Chinose, H.; Chye, T.; Fujita, K.; Yoshida, M.; Ueda, S.; Nagatsu, T.
Biochem. Biophys. Res. Commun. 195, 158-165, 1993
A:Title: Increased heterogeneity of tyrosine hydroxylase in humans.
A:Reference number: PN0575; MUID:93371398; PMID:7689834
A:Accession: PN0583
A:Molecule type: Genomic RNA
A:Residues: 1-27 <ICH>
A:Cross-references: GB:L14790
A:Experimental source: lymphocytes of peripheral blood
C:Comment: This enzyme catalyzes the first and rate-limiting step of catecholamine bio
C:Superfamily: phenylalanine 4-monooxygenase
C:Keywords: bipterin; monooxygenase; oxidoreductase

Query Match 100.0%; Score 21; DB 2; Length 27;
Best Local Similarity 100.0%; Pred. No. 1.7e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GPSL 4

Db 4 GPSSL 7

RESULT 3

PN0584
tyrosine 3-monooxygenase (EC 1.14.16.2) type 3 and 4 - gorilla (fragment)
N:Alternate names: oxygen oxidoreductase; tetrahydropteridine; tyrosine hydroxylase
C:Species: Gorilla gorilla (gorilla)
C>Date: 03-May-1994 #sequence_revision 07-Oct-1994 #text_change 31-Mar-2000
C:Accession: PN0584
R:Richinose, H.; Ohye, T.; Fujita, K.; Yoshida, M.; Ueda, S.; Nagatsu, T.
Biochem. Biophys. Res. Commun. 195, 158-165, 1993
A:Title: Increased heterogeneity of tyrosine hydroxylase in humans.
A:Reference number: PN0575; MUID:93371398; PMID:7689834
A:Accession: PN0584
A:Molecule type: genomic RNA
A:Residues: 1-27 <ICH>
A:Cross-references: GB:L14796
A:Experimental source: lymphocytes of peripheral blood
C:Comment: This enzyme catalyzes the first and rate-limiting step of catecholamine biosynthesis
C:Superfamily: phenylalanine 4-monooxygenase
C:Keywords: bipterin; monooxygenase; oxidoreductase

Query Match 100.0%; Score 21; DB 2; Length 27;

Best Local Similarity 100.0%; Pred. NO. 1.7e+02; Indels 0; Gaps 0;

Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GPSSL 4

Db 4 GPSSL 7

RESULT 4

PN0585
tyrosine 3-monooxygenase (EC 1.14.16.2) type 3 and 4 - orangutan (fragment)
N:Alternate names: oxygen oxidoreductase; tetrahydropteridine; tyrosine hydroxylase
C:Species: Pongo pygmaeus (orangutan)
C>Date: 03-May-1994 #sequence_revision 07-Oct-1994 #text_change 31-Mar-2000
C:Accession: PN0585
R:Richinose, H.; Ohye, T.; Fujita, K.; Yoshida, M.; Ueda, S.; Nagatsu, T.
Biochem. Biophys. Res. Commun. 195, 158-165, 1993
A:Title: Increased heterogeneity of tyrosine hydroxylase in humans.
A:Reference number: PN0575; MUID:93371398; PMID:7689834
A:Accession: PN0585
A:Molecule type: genomic RNA
A:Residues: 1-27 <ICH>
A:Cross-references: GB:L14799
A:Experimental source: lymphocytes of peripheral blood
C:Comment: This enzyme catalyzes the first and rate-limiting step of catecholamine biosynthesis
C:Superfamily: phenylalanine 4-monooxygenase
C:Keywords: bipterin; monooxygenase; oxidoreductase

Query Match 100.0%; Score 21; DB 2; Length 27;

Best Local Similarity 100.0%; Pred. NO. 1.7e+02; Indels 0; Gaps 0;

Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GPSSL 4

Db 4 GPSSL 7

RESULT 5

PN0586
tyrosine 3-monooxygenase (EC 1.14.16.2) type 3 and 4 - common gibbon (fragment)
N:Alternate names: oxygen oxidoreductase; tetrahydropteridine; tyrosine hydroxylase
C:Species: Hylobates lar (common gibbon, white-handed gibbon)
C>Date: 03-May-1994 #sequence_revision 07-Oct-1994 #text_change 31-Mar-2000
C:Accession: PN0586
R:Richinose, H.; Ohye, T.; Fujita, K.; Yoshida, M.; Ueda, S.; Nagatsu, T.
Biochem. Biophys. Res. Commun. 195, 158-165, 1993
A:Title: Increased heterogeneity of tyrosine hydroxylase in humans.
A:Reference number: PN0575; MUID:93371398; PMID:7689834

A:Accession: PN0586
A:Molecule type: genomic RNA
A:Residues: 1-27 <ICH>
A:Cross-references: GB:L14793
A:Experimental source: lymph nodes
C:Comment: This enzyme catalyzes the first and rate-limiting step of catecholamine biosynthesis
C:Superfamily: phenylalanine 4-monooxygenase
C:Keywords: bipterin; monooxygenase; oxidoreductase

Query Match 100.0%; Score 21; DB 2; Length 27;

Best Local Similarity 100.0%; Pred. NO. 1.7e+02; Indels 0; Gaps 0;

Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GPSSL 4

Db 4 GPSSL 7

RESULT 6

S20771
Ig heavy chain V region (VH4, INIP24) - human (fragment)
C:Species: Homo sapiens (man)
C>Date: 20-Feb-1995 #sequence_revision 19-Apr-1996 #text_change 23-Jul-1999
C:Accession: S20771
R:Mortari, F.; Wang, J.; Schroeder, H.W.
submitted to the EMBL Data Library, April 1992
A:Description: Analysis of human cord blood Ig heavy chain IGA and IGG repertoire.
A:Reference number: S20764
A:Accession: S20771
A:Molecule type: DNA
A:Residues: 1-32 <MOR>
A:Cross-references: EMBL:Z11955; NID:G33871; PIDN:CAA78012.1; PID:G33872
C:Superfamily: immunoglobulin V region; immunoglobulin homology
C:Keywords: heterotetramer; immunoglobulin

Query Match 100.0%; Score 21; DB 2; Length 32;

Best Local Similarity 100.0%; Pred. NO. 2e+02; Indels 0; Gaps 0;

Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GPSSL 4

Db 16 GPSSL 19

RESULT 7

C25159
hypothetical protein 3 4.2K sin - Bacillus subtilis
C:Species: Bacillus subtilis
C>Date: 16-Aug-1988 #sequence_revision 16-Aug-1988 #text_change 16-Dec-1998
C:Accession: C25159
R:Gaur, N.K.; Dabnau, E.; Smith, I.
J. Bacteriol. 158, 860-869, 1986
A:Title: Characterization of a cloned Bacillus subtilis gene that inhibits sporulation
A:Reference number: A91827; MUID:87056981; PMID:3096962
A:Accession: C25159
A:Molecule type: DNA
A:Residues: 1-35 <GAU>

Query Match 100.0%; Score 21; DB 2; Length 35;

Best Local Similarity 100.0%; Pred. NO. 2.1e+02; Indels 0; Gaps 0;

Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GPSSL 4

Db 4 GPSSL 7

RESULT 8

SW1QDA
diuretic hormone - African migratory locust
C:Species: Locusta migratoria migratoroides (African migratory locust)
C>Date: 31-Dec-1991 #sequence_revision 31-Dec-1991 #text_change 03-Nov-2003
C:Accession: JH0243; S17864; A23702

R;Lehmberg, E.; Ota, R.B.; Furuya, K.; King, D.S.; Applebaum, S.W.; Ferenz, H.J.; School Biochem. Biophys. Res. Commun. 179, 1036-1041, 1991
A;Title: Identification of a diuretic hormone of Locusta migratoria.
A;Reference number: JH0243; MUID:91378968; PMID:1654896
A;Accession: JH0243
A;Molecule type: protein
A;Residues: 1-46 <LE2>
A;Experimental source: brain and corpora cardiaca
R;Kay, I.; Wheeler, C.H.; Coast, G.M.; Totty, N.F.; Cusinato, O.; Patel, M.; Goldsworthy B.I.; Chem. Hoppe-Seyler 372, 929-934, 1991
A;Title: Characterization of a diuretic peptide from Locusta migratoria.
A;Reference number: S17864; MUID:92126231; PMID:1663363
A;Accession: S17864
A;Molecule type: protein
A;Residues: 1-46 <KAY>
A;Note: species designated as Locusta migratoria
C;Comment: This hormone stimulates urine production by Malpighian tubules and elevates diuresis.
C;Superfamily: diuretic hormone; diuretic hormone homology
F;2-46/Domain: amidated carboxyl end; diuretic; hormone; osmoregulation
F;46/Modified site: amidated carboxyl end (ile) #status experimental

Query Match 100.0%; Score 21; DB 1; Length 46;
Best Local Similarity 100.0%; Pred. No. 2.8e+02; Indels 0; Gaps 0;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GPSL 4
|||
Db 4 GPSL 7

RESULT 9
A48542
CRF-related diuretic peptide - American cockroach
C;Species: Periplaneta americana (American cockroach)
C;Date: 19-Nov-1993 #sequence_revision 18-Nov-1994 #text_change 03-Nov-2003
C;Accession: A48542
R;Kay, I.; Patel, M.; Coast, G.M.; Totty, N.F.; Mallet, A.I.; Goldsworthy, G.J. Regul. Pept. 42, 111-122, 1992
A;Title: Isolation, characterization and biological activity of a CRF-related diuretic peptide.
A;Reference number: A48542; MUID:93174045; PMID:1337794
A;Accession: A48542
A;Status: preliminary
A;Molecule type: protein
A;Residues: 1-46 <KAY>
A;Note: sequence extracted from NCBI backbone (NCBIP:125546)
C;Superfamily: diuretic hormone; diuretic hormone homology
F;2-46/Domain: diuretic hormone homology <DRH>

Query Match 100.0%; Score 21; DB 2; Length 46;
Best Local Similarity 100.0%; Pred. No. 2.8e+02; Indels 0; Gaps 0;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GPSL 4
|||
Db 4 GPSL 7

RESULT 10
F90584
hypochemical protein MYPU 5820 [imported] - Mycoplasma pulmonis (strain UAB CTIP)
C;Species: Mycoplasma pulmonis
C;Date: 24-May-2001 #sequence_revision 24-May-2001 #text_change 03-Aug-2001
C;Accession: F90584
R;Chambaud, I.; Heilig, R.; Ferris, S.; Barbe, V.; Samson, D.; Gallissot, F.; Moszer, I.; Nucleic Acids Res. 29, 2145-2153, 2001
A;Title: The complete genome sequence of the murine respiratory pathogen Mycoplasma pulmonis.
A;Reference number: A99512; MUID:21267165; PMID:11353084
A;Accession: F90584
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-50 <KUR>
A;Cross-references: GB:AL445566; PID:g14089997; PIDN:CAC13755.1; GSPDB:GN00153

A;Experimental source: strain UAB CTIP
C;Genetics:
A;Gene: MYPU 5820
A;Genetic code: SGC3

Query Match 100.0%; Score 21; DB 2; Length 50;
Best Local Similarity 100.0%; Pred. No. 3e+02; Indels 0; Gaps 0;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GPSL 4
|||
Db 41 GPSL 44

RESULT 11
C38230
inorganic diphosphatase (EC 3.6.1.1) 1 - beet (fragments)
C;Species: Beta vulgaris (beet)
C;Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 27-Oct-2003
C;Accession: C38230; F38230
R;Sarrafian, V.; Kim, Y.; Poole, R.J.; Rea, P.A. Proc. Natl. Acad. Sci. U.S.A. 89, 1775-1779, 1992
A;Title: Molecular cloning and sequence of cDNA encoding the pyrophosphate-energized ATPase.
A;Reference number: A38230; MUID:92179265; PMID:1311852
A;Accession: C38230
A;Molecule type: protein
A;Residues: 1-16 <SARI>
A;Note: sequence extracted from NCBI backbone (NCBIP:87192)
A;Accession: E38230
A;Molecule type: protein
A;Residues: 17-28 <SAR2>
A;Note: sequence extracted from NCBI backbone (NCBIP:87194)
A;Accession: F38230
A;Molecule type: protein
A;Residues: 29-52 <SAR3>
A;Experimental source: cultivar Detroit Dark
C;Superfamily: H(+)-translocating inorganic pyrophosphatase
C;Keywords: hydrolase

Query Match 100.0%; Score 21; DB 2; Length 52;
Best Local Similarity 100.0%; Pred. No. 3.1e+02; Indels 0; Gaps 0;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GPSL 4
|||
Db 45 GPSL 48

RESULT 12
S35628
myosin heavy chain - rabbit (fragment)
C;Species: Oryctolagus cuniculus (domestic rabbit)
C;Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 13-Aug-1999
C;Accession: S35628
R;Babij, P. Nucleic Acids Res. 21, 1467-1471, 1993
A;Title: Tissue-specific and developmentally regulated alternative splicing of a viscose myosin heavy chain.
A;Reference number: S35628; MUID:93219127; PMID:8464739
A;Accession: S35628
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-54 <BAB>
A;Cross-references: EMBL:X70964; NID:G297108; PIDN:CAA50299.1; PID:G939958
C;Superfamily: myosin heavy chain; myosin motor domain homology
F;1-54/Domain: myosin motor domain homology (fragment) <MMOT>

Query Match 100.0%; Score 21; DB 2; Length 54;
Best Local Similarity 100.0%; Pred. No. 3.2e+02; Indels 0; Gaps 0;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GPSL 4
|||

Db 23 GPSTL 26

RESULT 13

T31181
 Hypothetical protein S05 - Sphingomonas aromaticivorans plasmid pNLI
 C:Species: Sphingomonas aromaticivorans
 C:Date: 11-Jan-2000 #sequence_revision 11-Jan-2000 #text_change 11-Jan-2000
 R:Accession: T31181
 R:Romaine, M.F.; Stillwell, L.C.; Wong, K.K.; Thurston, S.J.; Sisk, E.C.; Sensen, C.W.; C
 submitted to the EMBL Data Library, July 1999
 A:Description: Complete sequence of a 184 kb catabolic plasmid from Sphingomonas aromati
 A:Reference number: 420992
 A:Accession: T31181
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-54 <ROM>
 A:Cross-references: EMBL:AF079317; NID:g3378261; PID:g3378322; PIDN:RAD03905.1
 C:Genetics:
 A:Genome: plasmid pNLI
 A:Note: ori505

Query Match 100.0%; Score 21; DB 2; Length 54;

Best Local Similarity 100.0%; Pred. No. 3.2e+02;

Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GPSTL 4

Db 33 GPSTL 36

RESULT 14

T03658
 phosphoenolpyruvate carboxylase (EC 4.1.1.31) - maize (fragment).
 C:Species: Zea mays (maize)
 C:Date: 24-Mar-1999 #sequence_revision 24-Mar-1999 #text_change 22-Jun-1999
 R:Accession: T03658
 R:Yanagisawa, S.; Izui, K.; Yanaguchi, Y.; Shigesada, K.; Katsuki, H.
 FEBS Lett. 229, 107-110, 1988
 A:Title: Further analysis of cDNA clones for maize phosphoenolpyruvate carboxylase invol
 mRNA at multiple sites in vivo.
 A:Reference number: S00348; MUID:88152202; PMID:2894322
 A:Accession: T03658
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1-56 <YAN>
 A:Cross-references: EMBL:X07171; NID:g22414; PIDN:CAA30159.1; PID:g9299919

C:Function:
 A:Description: catalyzes the irreversible beta-carboxylation of phosphoenolpyruvate by p
 A:Pathway: carbon dioxide fixation
 C:Superfamily: phosphoenolpyruvate carboxylase
 C:Keywords: carbon dioxide fixation; carbon-carbon lyase; carboxy-lyase; phosphoprotein

Query Match

100.0%; Score 21; DB 2; Length 56;

Best Local Similarity 100.0%; Pred. No. 3.4e+02;

Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GPSTL 4

Db 28 GPSTL 31

RESULT 15

S36380
 Ig heavy chain V region - mouse (fragment)
 C:Species: Mus musculus (house mouse)
 C:Date: 19-Mar-1997 #sequence_revision 29-Aug-1997 #text_change 23-Jul-1999
 R:Accession: S36380; S33395
 R:Ansell, K.H.
 submitted to the EMBL Data Library, April 1993
 A:Reference number: S36376
 A:Accession: S36380
 A:Molecule type: mRNA

A:Residues: 1-56 <ANS>
 A:Cross-references: EMBL:X73008; NID:g295875; PIDN:CAA51495.1; PID:g939938
 R:Kettleborough, C.A.; Saldanha, J.; Ansell, K.H.; Bendig, M.M.
 Eur. J. Immunol. 23, 206-211, 1993

A:Title: Optimization of primers for cloning libraries of mouse immunoglobulin genes us
 A:Reference number: S33391; MUID:93122092; PMID:8419173

A:Accession: S33395

A:Molecule type: mRNA

A:Residues: 5-45 <KET>

A:Cross-references: EMBL:X73008

A:Experimental source: strain BALB/c

C:Superfamily: immunoglobulin V region; immunoglobulin homology

C:Keywords: heterotrimer; immunoglobulin

F:18-56/Domain: immunoglobulin homology (fragment) <IMM>

Query Match 100.0%; Score 21; DB 2; Length 56;

Best Local Similarity 100.0%; Pred. No. 3.4e+02;

Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GPSTL 4

Db 12 GPSTL 15

Search completed: March 10, 2004, 09:16:43

Job time : 3.30396 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 10, 2004, 09:16:59 ; Search time 15.7198 Seconds
(without alignments)
268.645 Million cell updates/sec

Title: US-09-848-834A-3
Perfect score: 95
Sequence: 1 DEKXIAKXKASSVFNVNS 20

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 809742 seqs, 211153259 residues

Total number of hits satisfying chosen parameters: 809742

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications AA:*
1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep.*
10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep.*
11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep.*
14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep.*
15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. NO. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	95	100.0	20	9	US-09-848-834A-3
2	95	100.0	33	9	US-09-848-834A-12
3	95	100.0	36	9	US-09-848-834A-16
4	95	100.0	51	9	US-09-848-834A-20
5	89	93.7	19	14	US-10-239-313A-54
6	89	93.7	21	10	US-09-932-165-1482
7	89	93.7	21	10	US-09-935-384-710
8	89	93.7	21	10	US-09-942-052-711
9	89	93.7	21	14	US-10-001-469-1404
10	89	93.7	21	14	US-10-128-711-97
11	89	93.7	21	14	US-10-116-118-33
12	89	93.7	21	14	US-10-062-109A-761
13	89	93.7	21	14	US-10-005-480A-761
14	89	93.7	21	14	US-10-277-292-652
15	89	93.7	21	15	US-10-291-241-26

Sequence 652, Appl
Sequence 20, Appl
Sequence 2583, Appl
Sequence 2, Appl
Sequence 52, Appl
Sequence 2, Appl
Sequence 44, Appl
Sequence 4224, Appl
Sequence 33, Appl
Sequence 23, Appl
Sequence 15, Appl
Sequence 17, Appl
Sequence 39, Appl
Sequence 31, Appl
Sequence 95, Appl
Sequence 53, Appl
Sequence 3, Appl
Sequence 36, Appl
Sequence 38, Appl
Sequence 39, Appl
Sequence 41, Appl
Sequence 60, Appl
Sequence 64, Appl
Sequence 67, Appl
Sequence 59, Appl
Sequence 65, Appl
Sequence 66, Appl
Sequence 14, Appl
Sequence 25, Appl

US-10-280-340-652
US-10-099-460-20
US-10-024-552-2583
US-10-120-885A-25
US-10-107-532-2
US-10-121-016-52
US-10-114-669-2
US-10-120-835-44
US-10-149-138-4224
US-10-114-432-39
US-09-747-502-23
US-09-865-294-15
US-10-411-544-17
US-10-411-544-39
US-09-820-843A-31
US-10-128-711-96
US-10-239-313A-53
US-10-372-111-3
US-10-239-313A-36
US-10-239-313A-39
US-10-239-313A-41
US-10-239-313A-60
US-10-239-313A-64
US-10-239-313A-67
US-10-239-313A-59
US-10-239-313A-65
US-10-239-313A-66
US-10-239-313A-14
US-10-239-313A-26

ALIGNMENTS

RESULT 1
US-09-848-834A-3
; Sequence 3, Application US/09848834A
; Patent No. US20020076416A1
; GENERAL INFORMATION:
; APPLICANT: Aption Corporation
; TITLE OF INVENTION: Chimeric Peptide Immunogens
; FILE REFERENCE: 1102865-0047
; CURRENT APPLICATION NUMBER: US/09/848,834A
; CURRENT FILING DATE: 2001-05-04
; PRIOR APPLICATION NUMBER: 60/202,328
; PRIOR FILING DATE: 2000-05-05
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: Patent in version 3.0
; SEQ ID NO 3
; TYPE: PRT
; LENGTH: 20
; ORGANISM: Plasmodium falciparum
; FEATURE:
; NAME/KEY: PEPTIDE
; LOCATION: (1)..(20)
; OTHER INFORMATION: Amino acid sequence 378-398 of the Plamodium
; OTHER INFORMATION: falciparum circumsporozoite (CSP) protein
US-09-848-834A-3

Query Match 100.0%; Score 95; DB 9; Length 20;
Best Local Similarity 100.0%; Pred. NO. 2.9e-08;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 DEKXIAKXKASSVFNVNS 20
DB 1 DEKXIAKXKASSVFNVNS 20

RESULT 2
US-09-848-834A-12
; Sequence 12, Application US/09848834A
; Patent No. US20020076416A1


```

; GENERAL INFORMATION:
; APPLICANT: Apton Corporation
; TITLE OF INVENTION: Chimeric Peptide Immunogens
; FILE REFERENCE: 1102865-0047
; CURRENT APPLICATION NUMBER: US/09/848,834A
; CURRENT FILING DATE: 2001-05-04
; PRIOR APPLICATION NUMBER: 60/202,328
; PRIOR FILING DATE: 2000-05-05
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 12
; LENGTH: 33
; TYPE: PPT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Chimeric peptide consisting of amino acid sequence 378-398 of the
; OTHER INFORMATION: Plasmodium falciparum CSP protein linked by a spacer to amino ac
; OTHER INFORMATION: id sequence 2-10 of the GnRH hormone
; NAME/KEY: MOD_RES
; LOCATION: (1)..(1)
; OTHER INFORMATION: Amidated aspartic acid
; NAME/KEY: MOD_RES
; LOCATION: (33)..(33)
; OTHER INFORMATION: Amidated glycine or glycineamide
; NAME/KEY: PEPTIDE
; LOCATION: (1)..(20)
; OTHER INFORMATION: Amino acid sequence 378-398 of the Malaria
; OTHER INFORMATION: (Plasmodium falciparum) circumsporozoite
; NAME/KEY: PEPTIDE
; LOCATION: (21)..(24)
; OTHER INFORMATION: Spacer peptide
; NAME/KEY: PEPTIDE
; LOCATION: (25)..(33)
; OTHER INFORMATION: Amino acid sequence 2-10 of the human GnRH hormone
; US-09-848-834A-12

Query Match 100.0%; Score 95; DB 9; Length 33;
Best Local Similarity 100.0%; Pred. No. 5e-08;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DEKKIARMEKASSVFNVNS 20
DB 1 DEKKIARMEKASSVFNVNS 20

RESULT 3
US-09-848-834A-16
; Sequence 16, Application US/09848834A
; Patent No. US20020076416A1
; GENERAL INFORMATION:
; APPLICANT: Apton Corporation
; TITLE OF INVENTION: Chimeric Peptide Immunogens
; FILE REFERENCE: 1102865-0047
; CURRENT APPLICATION NUMBER: US/09/848,834A
; CURRENT FILING DATE: 2001-05-04
; PRIOR APPLICATION NUMBER: 60/202,328
; PRIOR FILING DATE: 2000-05-05
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 16
; LENGTH: 36
; TYPE: PPT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Chimeric peptide consisting of amino acid sequence 1-10 of the Gn
; OTHER INFORMATION: RH hormone linked by a spacer to amino acid sequence 378-398 of
; OTHER INFORMATION: the Plasmodium falciparum circumsporozoite (CSP) protein
; NAME/KEY: MOD_RES
; LOCATION: (1)..(1)
; OTHER INFORMATION: Pyroglutamic acid or 5-oxoproline
; NAME/KEY: PEPTIDE
; LOCATION: (1)..(10)

```

```

; OTHER INFORMATION: Amino acid sequence 1-10 of the human GnRH hormone
; NAME/KEY: PEPTIDE
; LOCATION: (11)..(16)
; OTHER INFORMATION: Spacer peptide
; NAME/KEY: PEPTIDE
; LOCATION: (17)..(36)
; OTHER INFORMATION: Amino acid sequence 378-398 of the Malaria
; OTHER INFORMATION: (Plasmodium falciparum) circumsporozoite
; OTHER INFORMATION: (CSP) protein
; US-09-848-834A-16

Query Match 100.0%; Score 95; DB 9; Length 36;
Best Local Similarity 100.0%; Pred. No. 5.5e-08;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DEKKIARMEKASSVFNVNS 20
DB 17 DEKKIARMEKASSVFNVNS 36

RESULT 4
US-09-848-834A-20
; Sequence 20, Application US/09848834A
; Patent No. US20020076416A1
; GENERAL INFORMATION:
; APPLICANT: Apton Corporation
; TITLE OF INVENTION: Chimeric Peptide Immunogens
; FILE REFERENCE: 1102865-0047
; CURRENT APPLICATION NUMBER: US/09/848,834A
; CURRENT FILING DATE: 2001-05-04
; PRIOR APPLICATION NUMBER: 60/202,328
; PRIOR FILING DATE: 2000-05-05
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 20
; LENGTH: 51
; TYPE: PPT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Chimeric peptide consisting of amino acid sequence 1-10 of human
; OTHER INFORMATION: GnRH linked by a spacer to amino acid sequence 378-398 of Plasm
; OTHER INFORMATION: ium falciparum circumsporozoite (CSP) protein
; NAME/KEY: MOD_RES
; LOCATION: (1)..(1)
; OTHER INFORMATION: Pyroglutamic acid or 5-oxoproline
; NAME/KEY: MOD_RES
; LOCATION: (51)..(51)
; OTHER INFORMATION: Amidated glycine or glycineamide
; NAME/KEY: PEPTIDE
; LOCATION: (1)..(10)
; OTHER INFORMATION: Amino acid sequence 1-10 of the human GnRH hormone
; NAME/KEY: PEPTIDE
; LOCATION: (11)..(16)
; OTHER INFORMATION: Spacer peptide
; NAME/KEY: PEPTIDE
; LOCATION: (17)..(36)
; OTHER INFORMATION: Amino acid sequence 378-398 of the Plasmodium falciparum
; OTHER INFORMATION: circumsporozoite (CSP) protein
; NAME/KEY: PEPTIDE
; LOCATION: (37)..(42)
; OTHER INFORMATION: Spacer peptide
; NAME/KEY: PEPTIDE
; LOCATION: (43)..(51)
; OTHER INFORMATION: Amino acid sequence 2-10 of the human GnRH hormone
; US-09-848-834A-20

Query Match 100.0%; Score 95; DB 9; Length 51;
Best Local Similarity 100.0%; Pred. No. 8e-08;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DEKKIARMEKASSVFNVNS 20
DB 17 DEKKIARMEKASSVFNVNS 36

```

RESULT 5
US-10-239-313A-54
; Sequence 54, Application US/10239313A
; Publication No. US20030175285A1
; GENERAL INFORMATION:
; APPLICANT: KLINGUER - HAMOUR, Christine
; APPLICANT: CORVAIA, Nathalie
; APPLICANT: BECK, Alain
; APPLICANT: GOETSCH, Liliane
; TITLE OF INVENTION: MOLECULE OF PHARMACEUTICAL INTEREST COMPRISING AT ITS
; TITLE OF INVENTION: N-TERMINAL A GLUTAMIC ACID OR A GLUTAMINE IN THE FORM
; TITLE OF INVENTION: OF A PHYSIOLOGICALLY ACCEPTABLE STRONG ACID
; FILE REFERENCE: 343 727 - US
; CURRENT APPLICATION NUMBER: US/10/239,313A
; PRIOR FILING DATE: 2002-09-19
; PRIOR FILING DATE: 2000-03-23
; PRIOR FILING DATE: 2000-03-23
; PRIOR FILING DATE: 2001-03-22
; NUMBER OF SEQ ID NOS: 697
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 54
; LENGTH: 19
; TYPE: PRT
; ORGANISM: Plasmodium malariae
US-10-239-313A-54

Query Match 93.7%; Score 89; DB 14; Length 19;
Best Local Similarity 100.0%; Pred. No. 2.3e-07;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 EKXIAKMEKASSVFNVNS 20
DB 1 EKXIAKMEKASSVFNVNS 19

RESULT 6
US-09-932-165-1482
; Sequence 1482, Application US/09932165
; Publication No. US20030134784A1
; GENERAL INFORMATION:
; APPLICANT: RAITANO, ARTHUR
; APPLICANT: CHALLITA-EID, PIA M.
; APPLICANT: FARIS, MARY
; APPLICANT: SAPPAN, DOUGLAS
; APPLICANT: AFAR, DANIEL
; APPLICANT: LEVIN, ELANA
; APPLICANT: HUBERT, RENE
; APPLICANT: GE, WANGMAO
; APPLICANT: JAKOBOWITS, AYA
; TITLE OF INVENTION: NUCLEIC ACIDS AND CORRESPONDING PROTEINS ENTITLED
; TITLE OF INVENTION: 83P2H3 AND CatP2E11 USEFUL IN TREATMENT AND
; TITLE OF INVENTION: DETECTION OF CANCER
; FILE REFERENCE: 51158-20014.00
; CURRENT APPLICATION NUMBER: US/09/932,165
; CURRENT FILING DATE: 2001-08-17
; PRIOR FILING DATE: 2000-08-17
; PRIOR FILING DATE: 2000-08-17
; NUMBER OF SEQ ID NOS: 1508
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1482
; LENGTH: 21
; TYPE: PRT
; ORGANISM: Plasmodium falciparum
US-09-932-165-1482

Query Match 93.7%; Score 89; DB 10; Length 21;
Best Local Similarity 100.0%; Pred. No. 2.6e-07;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 EKXIAKMEKASSVFNVNS 20

DB 3 EKXIAKMEKASSVFNVNS 21

RESULT 7
US-09-935-384-710
; Sequence 710, Application US/09935384
; Publication No. US20030166526A1
; GENERAL INFORMATION:
; APPLICANT: CHALLITA-EID, PIA
; APPLICANT: HUBERT, RENE
; APPLICANT: RAITANO, ARTHUR
; APPLICANT: AFAR, DANIEL
; APPLICANT: LEVIN, ELANA
; APPLICANT: FARIS, MARY
; APPLICANT: GE, WANGMAO
; APPLICANT: JAKOBOWITS, AYA
; TITLE OF INVENTION: NUCLEIC ACID AND CORRESPONDING PROTEIN NAMED 158P1H4
; TITLE OF INVENTION: USEFUL IN THE TREATMENT AND DETECTION OF BLADDER AND
; FILE REFERENCE: 51158-20033.00
; CURRENT APPLICATION NUMBER: US/09/935,384
; CURRENT FILING DATE: 2001-08-22
; PRIOR FILING DATE: 2000-08-22
; PRIOR FILING DATE: 2000-08-22
; NUMBER OF SEQ ID NOS: 783
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 710
; LENGTH: 21
; TYPE: PRT
; ORGANISM: Plasmodium falciparum
US-09-935-384-710

Query Match 93.7%; Score 89; DB 10; Length 21;
Best Local Similarity 100.0%; Pred. No. 2.6e-07;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 EKXIAKMEKASSVFNVNS 20
DB 3 EKXIAKMEKASSVFNVNS 21

RESULT 8
US-09-942-052-711
; Sequence 711, Application US/09942052
; Publication No. US20030170626A1
; GENERAL INFORMATION:
; APPLICANT: Raitano, Arthur B.
; APPLICANT: Faris, Mary
; APPLICANT: Hubert, Rene S.
; APPLICANT: Afar, Daniel
; APPLICANT: Ge, Wangmao
; APPLICANT: Challita-Eid, Pia M.
; TITLE OF INVENTION: NUCLEIC ACID AND CORRESPONDING PROTEIN ENTITLED 85P1B3
; TITLE OF INVENTION: USEFUL IN TREATMENT AND DETECTION OF CANCER
; FILE REFERENCE: 51158-20028.00
; CURRENT APPLICATION NUMBER: US/09/942,052
; CURRENT FILING DATE: 2001-08-28
; PRIOR FILING DATE: 2000-08-28
; PRIOR FILING DATE: 2000-08-28
; NUMBER OF SEQ ID NOS: 744
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 711
; LENGTH: 21
; TYPE: PRT
; ORGANISM: Plasmodium falciparum
US-09-942-052-711

Query Match 93.7%; Score 89; DB 10; Length 21;
Best Local Similarity 100.0%; Pred. No. 2.6e-07;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 EKXIAKMEKASSVFNVNS 20

Db 3 EKKIAMEKASSVFNVNVS 21
 |||||

RESULT 9

US-10-001-469-1404
 ; Sequence 1404, Application US/10001469
 ; Publication No. US20030091562A1
 ; GENERAL INFORMATION:
 ; APPLICANT: JAKOBOVITS, AYA
 ; APPLICANT: RAITANO, DANIEL
 ; APPLICANT: APAR, DANIEL
 ; APPLICANT: SAPFRAN, DOUGLAS
 ; APPLICANT: HUBERT, RENE
 ; APPLICANT: PARIS, MARY
 ; APPLICANT: CHALLITA-EID, PIA
 ; TITLE OF INVENTION: NUCLEIC ACID AND CORRESPONDING PROTEIN ENTITLED
 ; FILE OF INVENTION: 101P3A11 USEFUL IN TREATMENT AND DETECTION OF CANCER
 ; FILE REFERENCE: 51158-20024.20
 ; CURRENT APPLICATION NUMBER: US/10/001.469
 ; CURRENT FILING DATE: 2002-05-20
 ; PRIOR APPLICATION NUMBER: 60/157,902
 ; PRIOR FILING DATE: 1999-10-05
 ; PRIOR APPLICATION NUMBER: 60/291,118
 ; PRIOR FILING DATE: 2001-05-15
 ; PRIOR APPLICATION NUMBER: 09/680,728
 ; PRIOR FILING DATE: 2000-10-05
 ; NUMBER OF SEQ ID NOS: 2888
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 1404
 ; LENGTH: 21
 ; TYPE: PRT
 ; ORGANISM: Plasmodium falciparum
 US-10-001-469-1404

Query Match 93.7%; Score 89; DB 14; Length 21;
 Best Local Similarity 100.0%; Pred. No. 2.6e-07;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 EKKIAMEKASSVFNVNVS 20
 |||||
 Db 3 EKKIAMEKASSVFNVNVS 21

RESULT 10

US-10-128-711-97
 ; Sequence 97, Application US/10128711
 ; Publication No. US20030099534A1
 ; GENERAL INFORMATION:
 ; APPLICANT: VITIELLO, Maria A.
 ; APPLICANT: CHESTNUT, Robert W.
 ; APPLICANT: SETTE, Alessandro D.
 ; APPLICANT: CELIS, Eteban
 ; APPLICANT: GRAY, Howard
 ; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR ELICITING
 ; CTL IMMUNITY
 ; NUMBER OF SEQUENCES: 153
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Townsend and Townsend Khourie and Crew
 ; STREET: Stuart Street Tower, One Market Plaza
 ; CITY: San Francisco
 ; STATE: California
 ; COUNTRY: US
 ; ZIP: 94105-1493
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA: US/10/128,711
 ; FILING DATE: 22-Apr-2002
 ; CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/197,484
 ; FILING DATE: 16-FEB-1994
 ; APPLICATION NUMBER: US 07/935,811
 ; FILING DATE: 26-AUG-1992
 ; APPLICATION NUMBER: US 07/874,491
 ; FILING DATE: 27-APR-1992
 ; APPLICATION NUMBER: US 07/827,682
 ; FILING DATE: 29-JAN-1992
 ; APPLICATION NUMBER: US 07/749,568
 ; FILING DATE: 26-AUG-1991
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Parmelee, Steven W.
 ; REGISTRATION NUMBER: 31,990
 ; REFERENCE/DOCKET NUMBER: 14137-26-4
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (206) 467-9600
 ; TELEFAX: (206) 623-6793
 ; INFORMATION FOR SEQ ID NO: 97:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 21 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS: unknown
 ; TOPOLOGY: unknown
 ; MOLECULE TYPE: peptide
 ; FEATURE:
 ; NAME/KEY: Peptide
 ; LOCATION: 1..21
 ; OTHER INFORMATION: /note= "Malaria circumsporozoite
 ; 378-398"
 ; SEQUENCE DESCRIPTION: SEQ ID NO: 97:
 US-10-128-711-97

Query Match 93.7%; Score 89; DB 14; Length 21;
 Best Local Similarity 100.0%; Pred. No. 2.6e-07;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 EKKIAMEKASSVFNVNVS 20
 |||||
 Db 3 EKKIAMEKASSVFNVNVS 21

RESULT 11

US-10-116-118-33
 ; Sequence 33, Application US/10116118
 ; Publication No. US20030143672A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Tangri, Shabnam
 ; APPLICANT: Sette, Alessandro
 ; APPLICANT: Ishioka, Glenn
 ; APPLICANT: Fikes, John D.
 ; TITLE OF INVENTION: Heteroclitic Analogs and Related Methods
 ; FILE REFERENCE: 2060.0090003
 ; CURRENT APPLICATION NUMBER: US/10/116,118
 ; CURRENT FILING DATE: 2002-08-07
 ; PRIOR APPLICATION NUMBER: US 60/166,529
 ; PRIOR FILING DATE: 1999-11-18
 ; PRIOR APPLICATION NUMBER: US 60/239,008
 ; PRIOR FILING DATE: 2000-10-06
 ; NUMBER OF SEQ ID NOS: 53
 ; SOFTWARE: PatentIn version 3.1
 ; SEQ ID NO 33
 ; LENGTH: 21
 ; TYPE: PRT
 ; ORGANISM: Plasmodium falciparum
 US-10-116-118-33

Query Match 93.7%; Score 89; DB 14; Length 21;
 Best Local Similarity 100.0%; Pred. No. 2.6e-07;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 EKKIAMEKASSVFNVNVS 20
 |||||

Db 3 EKKIAKMEKASSVFNVNS 21

RESULT 12
US-10-062-109A-761
; Sequence 761, Application US/10062109A
; Publication No. US20030165505A1
; GENERAL INFORMATION:
; APPLICANT: Agensys
; APPLICANT: Challita-Eid, Pia M.
; APPLICANT: Raitano, Arthur B.
; APPLICANT: Paris, Mary
; APPLICANT: Hubert, Rene S.
; APPLICANT: Morrison, Karen Jane Meyrick
; APPLICANT: Jakobovits, Aya
; TITLE OF INVENTION: Nucleic Acid and Corresponding Protein
; TITLE OF INVENTION: Entitled 161P2F10B Useful in Treatment and Detection of
; TITLE OF INVENTION: Cancer
; FILE REFERENCE: 51158-20062.01
; CURRENT APPLICATION NUMBER: US/10/062,109A
; PRIOR FILING DATE: 2002-01-31
; PRIOR APPLICATION NUMBER: US 10/005,480
; PRIOR FILING DATE: 2001-11-07
; NUMBER OF SEQ ID NOS: 765
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 761
; LENGTH: 21
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-062-109A-761

Query Match 93.7%; Score 89; DB 14; Length 21;
Best Local Similarity 100.0%; Pred. No. 2.6e-07;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 EKKIAKMEKASSVFNVNS 20
Db 3 EKKIAKMEKASSVFNVNS 21

RESULT 13
US-10-005-480A-761
; Sequence 761, Application US/10005480A
; Publication No. US20030191073A1
; GENERAL INFORMATION:
; APPLICANT: Agensys
; APPLICANT: Challita-Eid, Pia M.
; APPLICANT: Raitano, Arthur B.
; APPLICANT: Paris, Mary
; APPLICANT: Hubert, Rene S.
; APPLICANT: Morrison, Karen Jane Meyrick
; APPLICANT: Jakobovits, Aya
; TITLE OF INVENTION: Nucleic Acid and Corresponding Protein
; TITLE OF INVENTION: Entitled 161P2F10B Useful in Treatment and Detection of
; TITLE OF INVENTION: Cancer
; FILE REFERENCE: 51158-20062.00
; CURRENT APPLICATION NUMBER: US/10/005,480A
; CURRENT FILING DATE: 2001-11-07
; NUMBER OF SEQ ID NOS: 765
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 761
; LENGTH: 21
; TYPE: PRT
; ORGANISM: Homo sapien
US-10-005-480A-761

Query Match 93.7%; Score 89; DB 14; Length 21;
Best Local Similarity 100.0%; Pred. No. 2.6e-07;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 EKKIAKMEKASSVFNVNS 20
Db 3 EKKIAKMEKASSVFNVNS 21

RESULT 14
US-10-277-292-652
; Sequence 652, Application US/10277292
; Publication No. US20030199470A1
; GENERAL INFORMATION:
; APPLICANT: Farris, Mary
; APPLICANT: Hubert, Rene
; APPLICANT: Raitano, Arthur
; APPLICANT: AFAR, DANIEL
; APPLICANT: LEVIN, ELANA
; APPLICANT: CHALLITA-EID, PIA
; APPLICANT: JAKOBOVITS, AYA
; TITLE OF INVENTION: NUCLEIC ACID AND CORRESPONDING PROTEIN NAMED 158P1D7
; TITLE OF INVENTION: USEFUL IN THE TREATMENT AND DETECTION OF BLADDER AND
; TITLE OF INVENTION: OTHER CANCERS
; FILE REFERENCE: 51158-20050.00
; CURRENT APPLICATION NUMBER: US/10/277,292
; CURRENT FILING DATE: 2002-10-21
; PRIOR APPLICATION NUMBER: US/09/935,430
; PRIOR FILING DATE: 2001-08-22
; PRIOR APPLICATION NUMBER: 60/227,098
; PRIOR FILING DATE: 2000-08-22
; PRIOR APPLICATION NUMBER: 60/282,739
; PRIOR FILING DATE: 2001-04-10
; NUMBER OF SEQ ID NOS: 700
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 652
; LENGTH: 21
; TYPE: PRT
; ORGANISM: Plasmodium falciparum
US-10-277-292-652

Query Match 93.7%; Score 89; DB 14; Length 21;
Best Local Similarity 100.0%; Pred. No. 2.6e-07;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 EKKIAKMEKASSVFNVNS 20
Db 3 EKKIAKMEKASSVFNVNS 21

RESULT 15
US-10-291-241-26
; Sequence 26, Application US/10291241
; Publication No. US20030206905A1
; GENERAL INFORMATION:
; APPLICANT: Agensys, Inc.
; APPLICANT: Aya Jakobovits
; APPLICANT: Arthur B. Raitano
; APPLICANT: Mary Paris
; APPLICANT: Rene S. Hubert
; APPLICANT: Wangmao Ge
; APPLICANT: Karen Jane Meyrick Morrison
; APPLICANT: Robert Kendall Morrison
; APPLICANT: Pia M. Challita-Eid
; TITLE OF INVENTION: NUCLEIC ACID AND CORRESPONDING PROTEIN
; TITLE OF INVENTION: ENTITLED 161P2F10B USEFUL IN TREATMENT AND DETECTION OF
; FILE REFERENCE: 51158-20062.20
; CURRENT APPLICATION NUMBER: US/10/291,241
; CURRENT FILING DATE: 2003-03-18
; PRIOR APPLICATION NUMBER: US 10/005,480
; PRIOR FILING DATE: 2001-11-07
; PRIOR APPLICATION NUMBER: US 10/062,109
; PRIOR FILING DATE: 2002-01-31
; PRIOR APPLICATION NUMBER: US 60/282,739
; PRIOR FILING DATE: 2001-04-10
; NUMBER OF SEQ ID NOS: 103
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 26
; LENGTH: 21

us-09-848-834a-3.open.rapb

Wed Mar 10 10:34:33 2004

```

; TYPE: PRT
; ORGANISM: Plasmodium falciparum
US-10-291-241-26
Query Match      93.7%; Score 89; DB 15; Length 21;
Best Local Similarity 100.0%; Pred. No. 2.6e-07;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      2 EKXIARQEKASSVFNVNS 20
      |||||||
Db      3 EKXIARQEKASSVFNVNS 21
      |||||||

```

Search completed: March 10, 2004, 10:25:47
 Job time : 15.7198 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 10, 2004, 08:58:54 ; Search time 6.22568 Seconds
(without alignments)
309.015 Million cell updates/sec

Title: US-09-848-834A-3
Perfect score: 95
Sequence: 1 DEKKIAKMEKASSVFNVNS 20

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR 78.*

1: PIR1.*

2: PIR2.*

3: PIR3.*

4: PIR4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	81	85.3	388	2 A39756	circumsporozoite p
2	81	85.3	405	2 S05428	circumsporozoite p
3	81	85.3	412	1 QZQAF	circumsporozoite p
4	81	85.3	424	2 A54533	circumsporozoite p
5	81	85.3	442	2 A54529	circumsporozoite p
6	55	57.9	264	2 A44969	circumsporozoite p
7	55	57.9	332	1 QZQWB	circumsporozoite p
8	55	57.9	348	1 QZQBK	circumsporozoite p
9	55	57.9	367	1 QZQMI	circumsporozoite p
10	51	53.7	429	2 A54504	circumsporozoite p
11	51	53.7	485	2 A60610	circumsporozoite p
12	46	48.4	1534	2 A56734	circumsporozoite p
13	45	47.4	378	1 QZQAB	ribosome receptor,
14	45	47.4	378	1 QZQAL	circumsporozoite p
15	45	47.4	398	1 QZQAS	circumsporozoite p
16	45	47.4	401	1 QZQAC	circumsporozoite p
17	45	47.4	419	1 QZQAM	circumsporozoite p
18	44	46.3	255	2 E81265	probable periplasm
19	44	46.3	441	1 A45345	gag polyprotein 112 -
20	44	46.3	482	2 JCS593	amphiphysin II
21	44	46.3	529	2 A90328	hypothetical prote
22	44	46.3	994	2 S00960	hypothetical prote
23	43.5	45.8	388	2 JG6164	circumsporozoite p
24	43	45.3	280	2 H90556	fructose-bisphosph
25	43	45.3	343	2 A29319	circumsporozoite p
26	43	45.3	367	2 A32068	circumsporozoite p
27	43	45.3	378	1 QZQAV	circumsporozoite p
28	43	45.3	386	2 A48571	circumsporozoite p
29	43	45.3	387	2 D41156	circumsporozoite p

30 43 45.3 387 2 C41156 circumsporozoite p
31 43 45.3 395 2 A41156 circumsporozoite p
32 43 45.3 525 2 T40088 RhoGEF domain cont
33 43 45.3 627 2 T32958 hypothetical prote
34 42 44.2 175 2 H64701 membrane-associate
35 42 44.2 333 2 D64317 hypothetical prote
36 42 44.2 413 2 S59650 hypothetical prote
37 42 44.2 444 2 A11121 PTS betaglucoside-
38 42 44.2 444 2 A11481 PTS betaglucoside-
39 42 44.2 552 2 D92421 conserved hypothet
40 42 44.2 563 2 T40705 conserved hypothet
41 42 44.2 708 2 T50064 probable transcrip
42 42 44.2 773 2 H81362 phenylalanine-tRNA
43 42 44.2 921 2 D86293 F7H2.22 protein -
44 41.5 43.7 442 2 B82899 conserved hypothet
45 41 43.2 201 2 A31484 troponin I, fast s

ALIGNMENTS

RESULT 1

A39756 circumsporozoite protein - Plasmodium reichenowi

C/Species: Plasmodium reichenowi

C/Date: 14-Feb-1992 #sequence_revision 14-Feb-1992 #text_change 20-Aug-1999

C/Accession: A39756

R/Lal, A.A.; Goldman, I.F.

J. Biol. Chem. 266, 6686-6689, 1991

A/Title: Circumsporozoite protein gene from Plasmodium reichenowi, a chimpanzee malar
A/Reference number: A39756; MUID:91201303; PMID:2016283

A/Accession: A39756

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-388 <LAL>

A/Cross-references: GB:M60972; NID:g160228; PIDN:AAA29551.1; PID:g160229

C/Suprafamily: circumsporozoite protein; thrombospondin type 1 repeat homology
F;312-366/Domain: thrombospondin type 1 repeat homology <THR1>

Query Match 85.3%; Score 81; DB 2; Length 388;

Best Local Similarity 89.5%; Pred. No. 4.7e-05;

Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 EKKIAMEKASSVFNVNS 20

Db 356 EKKIAMEKASSVFNVNS 374

RESULT 2

S05428 circumsporozoite protein - malaria parasite (Plasmodium falciparum) (isolate NF54)

C/Species: Plasmodium falciparum

C/Date: 07-Jun-1990 #sequence_revision 07-Jun-1990 #text_change 09-Jun-2000

C/Accession: S05428; A45527; I60657

R/Campbell, J.R.

Nucleic Acids Res. 17, 5854, 1989

A/Title: DNA sequence of the gene encoding a Plasmodium falciparum malaria candidate v

A/Reference number: S05428; MUID:89345189; PMID:2668895

A/Accession: S05428

A/Status: translation not shown

A/Molecule type: DNA

A/Residues: 1-405 <CAM>

A/Cross-references: EMBL:X15363

R/Caspers, P.; Gents, R.; Matile, H.; Pink, J.R.; Sinigaglia, F.

Mol. Biochem. Parasitol. 35, 185-190, 1989

A/Title: The circumsporozoite protein gene from NF54, a Plasmodium falciparum isolate

A/Reference number: A45527; MUID:89364998; PMID:2671723

A/Accession: A45527

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-405 <CAS>

A/Cross-references: GB:M22982; GB:J04650; NID:g160168; PIDN:AAA29527.1; PID:g160169

R/Lockyer, M.J.; Marsh, K.; Newbold, C.I.

```

Mol. Biochem. Parasitol. 37, 275-280, 1989
A>Title: Wild isolates of Plasmodium falciparum show extensive polymorphism in T cell ep
A/Reference number: A60657; MUID:90114334; PMID:2481827
A/Accession: I60657
A>Status: Preliminary; not compared with conceptual translation
A/Molecule type: DNA
A/Residues: 319-336,354-373 <LOC>
A/Superfamily: circumsporozoite protein; thrombospondin type 1 repeat homology
F:329-383/Domain: thrombospondin type 1 repeat homology <THR1>

Query Match      85.3%; Score 81; DB 2; Length 408;
Best Local Similarity 89.5%; Pred. No. 4.9e-05;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 EKKIARKEKASSVFNVNS 20
Db 373 EKKICKMEKCSSVFNVNS 391

RESULT 3
OZQZAF
Circumsporozoite protein - malaria parasite (Plasmodium falciparum) (isolate IMTW22)
C/Species: Plasmodium falciparum
C/Date: 15-Nov-1984 #sequence_revision 15-Nov-1984 #text_change 09-Jun-2000
C/Accession: A03388
R/Dame, J.B.; Williams, J.L.; McCutchan, T.F.; Weber, J.L.; Wirtz, R.A.; Hochmeyer, W.T.
Science 225, 593-599, 1984
A>Title: Structure of the gene encoding the immunodominant surface antigen on the sporoz
A/Reference number: A03388; MUID:84250215; PMID:6204383
A/Accession: A03388
A/Molecule type: DNA
A/Residues: 1-412 <DAM>
A/Cross-references: GB:K02194; NID:G160160; PIDN:AAA29524.1; PID:G160161
A/Experimental source: clone 7G8
C/Comment: Residues 1-16 are the probable signal sequence.
C/Comment: There are 41 copies of a 4-residue repeating unit in the middle domain of the
C/Superfamily: circumsporozoite protein; thrombospondin type 1 repeat homology
F:336-390/Domain: thrombospondin type 1 repeat homology <THR1>

Query Match      85.3%; Score 81; DB 1; Length 412;
Best Local Similarity 89.5%; Pred. No. 5e-05;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 EKKIARKEKASSVFNVNS 20
Db 380 EKKICKMEKCSSVFNVNS 398

RESULT 4
A54533
Circumsporozoite protein - malaria parasite (Plasmodium falciparum) (strain T4, Thailand)
C/Species: Plasmodium falciparum
C/Date: 28-Oct-1994 #sequence_revision 28-Oct-1994 #text_change 09-Jun-2000
C/Accession: A54533
R/del Portillo, H.A.; Nussenzweig, R.S.; Enea, V.
Mol. Biochem. Parasitol. 24, 289-294, 1987
A>Title: Circumsporozoite gene of a Plasmodium falciparum strain from Thailand.
A/Reference number: A54533; MUID:87315205; PMID:3306373
A/Accession: A54533
A>Status: preliminary
A/Molecule type: DNA
A/Residues: 1-424 <DEL>
A/Cross-references: GB:M19752; NID:G160216; PIDN:AAA29555.1; PID:G160217
C/Superfamily: circumsporozoite protein; thrombospondin type 1 repeat homology
F:348-402/Domain: thrombospondin type 1 repeat homology <THR1>

Query Match      85.3%; Score 81; DB 2; Length 424;
Best Local Similarity 89.5%; Pred. No. 5.1e-05;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 EKKIARKEKASSVFNVNS 20
Db 392 EKKICKMEKCSSVFNVNS 410

```

RESULT 5

```

A54529
Circumsporozoite protein - malaria parasite (Plasmodium falciparum) (strain Wellcome)
C/Species: Plasmodium falciparum
C/Date: 28-Oct-1994 #sequence_revision 28-Oct-1994 #text_change 09-Jun-2000
C/Accession: A54529
R/Jockyer, M.J.; Schwarz, R.T.
Mol. Biochem. Parasitol. 22, 101-108, 1987
A>Title: Strain variation in the circumsporozoite protein gene of Plasmodium falciparum
A/Reference number: A54529; MUID:87115616; PMID:3543671
A/Accession: A54529
A>Status: preliminary; not compared with conceptual translation
A/Molecule type: DNA
A/Residues: 1-442 <LOC>
A/Cross-references: GB:M15505; NID:G160214; PIDN:AAA29554.1; PID:G160215
C/Superfamily: circumsporozoite protein; thrombospondin type 1 repeat homology
C/Keywords: tandem repeat
F:366-420/Domain: thrombospondin type 1 repeat homology <THR1>

```

```

Query Match      85.3%; Score 81; DB 2; Length 442;
Best Local Similarity 89.5%; Pred. No. 5.4e-05;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

```

QY 2 EKKIARKEKASSVFNVNS 20
Db 410 EKKICKMEKCSSVFNVNS 428

```

RESULT 6

```

A44969
Circumsporozoite protein precursor - Plasmodium yoelii nigeriensis
C/Species: Plasmodium yoelii nigeriensis
C/Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 01-Dec-1995
C/Accession: A44969
R/Colomer-Gould, V.; Enea, V.
Mol. Biochem. Parasitol. 43, 51-58, 1990
A>Title: Plasmodium yoelii nigeriensis circumsporozoite gene structure and its implicat
A/Reference number: A44969; MUID:91148645; PMID:2290446
A/Accession: A44969
A>Status: preliminary
A/Molecule type: DNA
A/Residues: 1-264 <COL>
A/Cross-references: GB:M32350
C/Superfamily: circumsporozoite protein; thrombospondin type 1 repeat homology
F:190-242/Domain: thrombospondin type 1 repeat homology <THR1>

```

```

Query Match      57.9%; Score 55; DB 2; Length 264;
Best Local Similarity 47.4%; Pred. No. 0.4;
Matches 9; Conservative 7; Mismatches 3; Indels 0; Gaps 0;

```

```

QY 2 EKKIARKEKASSVFNVNS 20
Db 232 DTEICKMDKCSIFNIVSN 250

```

RESULT 7

```

OZQZBE
Circumsporozoite protein precursor - Plasmodium berghei (strain NK65)
N/Alternate names: sporozoite surface antigen
C/Species: Plasmodium berghei
C/Date: 30-Sep-1987 #sequence_revision 28-Jul-1995 #text_change 16-Jul-1999
C/Accession: A44948; A25083; S13446
R/Lanar, D.E.
Mol. Biochem. Parasitol. 39, 151-154, 1990
A>Title: Sequence of the circumsporozoite gene of plasmodium berghei ANKA clone and NK6
A/Reference number: A44948; MUID:90158693; PMID:2406593
A/Accession: A44948
A/Molecule type: DNA
A/Residues: 1-332 <LAN>
A/Cross-references: GB:M28887
R/Eichinger, D.J.; Arnot, D.E.; Tam, J.P.; Nussenzweig, V.; Enea, V.

```

RESULT 9
OZZQMY
circumsporozoite protein precursor - Plasmodium yoelii
N/Alternate names: sporozoite surface antigen
C/Species: Plasmodium yoelii
C/Date: 30-Sep-1987 #sequence_revision 30-Sep-1987 #text_change 16-Jul-1999
C/Accession: A26271
R/Lal, A.A.; de la Cruz, V.F.; Welsh, J.A.; Charoenvit, Y.; Maloy, W.L.; McCutchan, T.
J Biol. Chem. 262, 2937-2940, 1987
A/Title: Structure of the gene encoding the circumsporozoite protein of Plasmodium yoelii
A/Reference number: A26271; MOID:87I37555; PMID:3102479
A/Accession: A26271
A/Molecule type: DNA
A/Residues: 1-367 xLAL>
A/Cross-references: GB:J02695; NID:g160222; PIDN:AAA29558.1; PID:g160223
C/Comment: There are three distinct regions in the mature circumsporozoite protein, the obic membrans-anchoring sequence.
C/Comment: There are 15 copies of a 6-residue repeat and 8 copies of a 4-residue repeat
C/Superfamily: circumsporozoite protein; thrombospondin type 1 repeat homology
C/Keywords: sporozoite; surface antigen; tandem repeat
F/1-19/Domain: signal sequence #status predicted <SIG>
F/20-367/Product: circumsporozoite protein #status predicted <MAT>
F/139-228/Region: 6-residue repeats (Q-G-P-G-A-P)
F/229-260/Region: 4-residue repeats (Q-Q-P-P)
F/293-345/Domain: thrombospondin type 1 repeat homology <THR1>

Query Match 57.9%; Score 55; DB 1; Length 367;
Best Local Similarity 47.4%; Pred. No. 0.56;
Matches 9; Conservative 7; Mismatches 3; Indels

QY 2 EKKIAMEKASSVFNVNS 20
: : | | | : | : | :
DB 335 DTEICKMDKCSSIFNIVSN 353

RESULT 10
A54504
Circumsporozoite protein - Plasmodium malariae (strain Uganda 1-CDC)
C.Species: Plasmodium malariae
C.Date: 06-Oct-1994 #sequence_revision 20-Aug-1999
C.Date: 06-Oct-1994 #sequence_revision 04-Nov-1994 #text_change 20-Aug-1999
C.Accession: A54504
L.Réal, A.A.; de la Cruz, V.F.; Campbell, G.H.; Procell, P.M.; Collins, W.E.; McCutchan
Mol. Biochem. Parasitol. 30, 291-294, 1988
A.Title: Structure of the circumsporozoite gene of plasmodium malariae.
A.Reference number: A54504 MUID:83040027; PMID:3054537

A:Molecule type: DNA
A:Residues: 1-429 <IALL>
A:/cross-references: GB:J03992; NID:g160220; PTDN:AAA29557.1; PT:
C:Superfamily: circumsporozoite protein; thrombospondin type 1
C/Keywords: tandem repeat
F:354-407/Domain: thrombospondin type 1 repeat homology <THR1>

```

Query Match          53.7% ; Score Nl; DB 2; Length 429;
Best Local Similarity 47.4%; Pred No. 2.8;
Matches 5; Conservative 6; Mismatches 4; Indels

Qy      : 2 EKXIAKHXEASVFNVNVS 20
         |::|::|::|::|::|::|::|
Db       397 ETEICSLDXCSIFNVNS 415
```

circumsporozoite protein precursor - Plasmodium brasilianum
C:Species: Plasmodium brasilianum
C:Date: 17-Apr-1993 #sequence revision 17-Apr-1993 #text change 20-Aug-1999

R. di Giovanni, L.; Cochrane, A.H.; Enea, V.
Exp. Parasitol. 70, 373-381, 1990
A. Title: On the evolutionary history of the circumsporozoite protein of *Plasmodium falciparum*
A. Reference number: A60510; PMID: 2323391

RESULT 13
OZZQAB
circumsporozoite protein precursor - Plasmodium cynomolgi (strain Berok)
N:Alternate names: major sporozoite surface antigen
C:Species: Plasmodium cynomolgi
C:Date: 30-Sep-1987 #sequence_revision 30-Sep-1987 #text_change 07-Nov-1997
C:Accession: D26255
R:Galinski, M.R.; Arnot, D.E.; Cochran, A.H.; Nussenzweig, R.S.; Enea, Cell 48, 311-319, 1987
A:Title: The circumsporozoite gene of the Plasmodium cynomolgi complex.
A:Reference number: A30889; MUID:87102878; PMID:3802196
A:Accession: D26255
A:Molecule type: DNA
A:Residues: 1-378 <GAL>
C:Comment: There are three distinct regions in the mature circumsporozoite protein, the obic membrane-anchoring sequence.
C:Comment: There are 10 tandem copies of a 9-residue repeat (preceded by a 6-residue in

RESULT 15
OZZQAS
circumsporozoite protein precursor - Plasmodium cynomolgi (strain Ceylon)
N:Alternate names: major sporozoite surface antigen
C:Species: Plasmodium cynomolgi
C:Date: 30-Sep-1987 #sequence_revision 30-Sep-1987 #text_change 07-Nov-1997
C:Accession: C36255
R:Galinski, M.R.; Arnot, D.E.; Cochran, A.H.; Barnwell, J.W.; Nussenzweig, R.S.; Enea
Cell 48, 311-319, 1987
A:Title: The circumsporozoite gene of the plasmodium cynomolgi complex.
A:Reference number: A30839; MUID:87102878; PMID:3802196
A:Accession: C36255
A:Molecule type: DNA
A:Residues: 1-398 <GAL>
C:Comment: There are three distinct regions in the mature circumsporozoite protein, the
oocyst membrane-anchoring sequence.
C:Comment: There are 16 tandem copies of a 9-residue repeat and 3 copies of a 17-residue
C:Superfamily: circumsporozoite protein; thrombospondin type 1 repeat homology
C:Keywords: sporozoite; surface antigen; tandem repeat
F:1-19/Domain: signal sequence #status predicted <SIG>
P:20-398/Product: circumsporozoite protein #status predicted <MAT>
F:97-240/Region: 9-residue repeats
F:97-240/Region: 9-residue repeats

```
; SEQUENCE CHARACTERISTICS:
; LENGTH: 10 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-019-010-2
Query Match          96.3%; Score 52; DB 9; Length 10;
Best Local Similarity 100.0%; Pred. No. 0.022;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 HWSYGLRP 9
Db 2 HWSYGLRP 9

RESULT 2
US-09-848-834A-1
; Sequence 1, Application US/09848834A
; Patent No. US20020076416A1
; GENERAL INFORMATION:
; APPLICANT: Apton Corporation
; TITLE OF INVENTION: Chimeric Peptide Immunogens
; FILE REFERENCE: 1102865-0047
; CURRENT APPLICATION NUMBER: US/09/848,834A
; CURRENT FILING DATE: 2001-05-04
; PRIOR APPLICATION NUMBER: 60/202,328
; PRIOR FILING DATE: 2000-05-05
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: Patentin version 3.0
; SEQ ID NO 1
; LENGTH: 10
; TYPE: PPT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: MOD_RES
; LOCATION: (1)..(1)
; OTHER INFORMATION: Pyroglutamic acid or 5-oxoproline
; NAME/KEY: MOD_RES
; LOCATION: (10)..(10)
; OTHER INFORMATION: Amidated-glycine or glycynamide
US-09-848-834A-1
Query Match          96.3%; Score 52; DB 9; Length 10;
Best Local Similarity 100.0%; Pred. No. 0.022;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 HWSYGLRP 9
Db 2 HWSYGLRP 9

RESULT 3
US-09-810-601-1
; Sequence 1, Application US/09810601
; Patent No. US20020177545A1
; GENERAL INFORMATION:
; APPLICANT: Donovan, Stephen
; TITLE OF INVENTION: Compositions and Methods for Treating Gonadotrophin
; FILE REFERENCE: 2947
; CURRENT APPLICATION NUMBER: US/09/810,601
; CURRENT FILING DATE: 2001-03-15
; NUMBER OF SEQ ID NOS: 44
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 1
; LENGTH: 10
; TYPE: PPT
; ORGANISM: human
; FEATURE:
; NAME/KEY: MOD_RES
; LOCATION: (1)
; OTHER INFORMATION: Xaa at position 1 is PyroGlu; PYRROLIDONE

US-09-810-601-1
Query Match          96.3%; Score 52; DB 9; Length 10;
Best Local Similarity 100.0%; Pred. No. 0.022;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 HWSYGLRP 9
Db 2 HWSYGLRP 9

RESULT 4
US-09-810-601-2
; Sequence 2, Application US/09810601
; Patent No. US20020177545A1
; GENERAL INFORMATION:
; APPLICANT: Donovan, Stephen
; TITLE OF INVENTION: Compositions and Methods for Treating Gonadotrophin
; FILE REFERENCE: 2947
; CURRENT APPLICATION NUMBER: US/09/810,601
; CURRENT FILING DATE: 2001-03-15
; NUMBER OF SEQ ID NOS: 44
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 2
; LENGTH: 10
; TYPE: PPT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Gonadotrophin
; OTHER INFORMATION: Release Hormone Analogue
; NAME/KEY: MOD_RES
; LOCATION: (1)
; OTHER INFORMATION: Xaa at position 1 is PyroGlu; PYRROLIDONE
; NAME/KEY: MOD_RES
; LOCATION: (10)
; OTHER INFORMATION: Xaa at position 10 is Gly-NH2; AMIDATION;
; PUBLICATION INFORMATION:
; PATENT DOCUMENT NUMBER: 5631226
; PATENT FILING DATE: 1993-07-07
; PUBLICATION DATE: 1997-05-20
; PUBLICATION INFORMATION:
; AUTHORS: Rousselle, Christophe
; JOURNAL: Mol. Pharmacol.
; VOLUME: 57
; PAGES: 679-686
; DATE: 2000
US-09-810-601-2
Query Match          96.3%; Score 52; DB 9; Length 10;
Best Local Similarity 100.0%; Pred. No. 0.022;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 HWSYGLRP 9
Db 2 HWSYGLRP 9

RESULT 5
US-09-810-601-3
; Sequence 3, Application US/09810601
; Patent No. US20020177545A1
; GENERAL INFORMATION:
; APPLICANT: Donovan, Stephen
; TITLE OF INVENTION: Compositions and Methods for Treating Gonadotrophin
; FILE REFERENCE: 2947
; CURRENT APPLICATION NUMBER: US/09/810,601
; CURRENT FILING DATE: 2001-03-15
; NUMBER OF SEQ ID NOS: 44
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 1
; LENGTH: 10
; TYPE: PPT
; ORGANISM: human
; FEATURE:
; NAME/KEY: MOD_RES
; LOCATION: (1)
; OTHER INFORMATION: Xaa at position 1 is PyroGlu; PYRROLIDONE
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.
OM protein - protein search, using sw model
Run on: March 10, 2004, 09:16:59 ; Search time 7.85992 Seconds
(without alignments)
268.645 Million cell updates/sec

Title: US-09-848-834A-1
Perfect score: 54
Sequence: 1 XHWSYGLRFX 10
Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5
Searched: 809742 seqs, 211153259 residues
Total number of hits satisfying chosen parameters: 809742

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications AA:
1: /cgn2_6/ptodata/2/pubaa/us07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/2/pubaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/2/pubaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/2/pubaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/2/pubaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/2/pubaa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/2/pubaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/2/pubaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/2/pubaa/US09A_PUBCOMB.pep.*
10: /cgn2_6/ptodata/2/pubaa/US09B_PUBCOMB.pep.*
11: /cgn2_6/ptodata/2/pubaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/2/pubaa/US09_NEW_PUB.pep.*
13: /cgn2_6/ptodata/2/pubaa/US10A_PUBCOMB.pep.*
14: /cgn2_6/ptodata/2/pubaa/US10B_PUBCOMB.pep.*
15: /cgn2_6/ptodata/2/pubaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/2/pubaa/US10_NEW_PUB.pep.*
17: /cgn2_6/ptodata/2/pubaa/US60_NEW_PUB.pep.*
18: /cgn2_6/ptodata/2/pubaa/US60_PUBCOMB.pep.*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES				%		Description	
Result No.	Score	Match	Length	DB	ID		
1	52	96.3	10	9	US-09-019-010-2	Sequence 2, Appli	
2	52	96.3	10	9	US-09-848-834A-1	Sequence 1, Appli	
3	52	96.3	10	9	US-09-810-601-1	Sequence 1, Appli	
4	52	96.3	10	9	US-09-810-601-2	Sequence 2, Appli	
5	52	96.3	10	9	US-09-810-601-3	Sequence 3, Appli	
6	52	96.3	10	9	US-09-810-601-4	Sequence 4, Appli	
7	52	96.3	10	10	US-09-964-201A-28	Sequence 28, Appli	
8	52	96.3	10	10	US-09-964-201A-32	Sequence 32, Appli	
9	52	96.3	10	10	US-09-305-924-1	Sequence 1, Appli	
10	52	96.3	10	10	US-09-305-924-9	Sequence 9, Appli	
11	52	96.3	10	13	US-10-109-331-2	Sequence 2, Appli	
12	52	96.3	10	13	US-10-184-126-1	Sequence 1, Appli	
13	52	96.3	10	14	US-10-115-553-1	Sequence 1, Appli	
14	52	96.3	10	14	US-10-122-483-1	Sequence 1, Appli	
15	52	96.3	10	14	US-10-117-364-1	Sequence 1, Appli	

16	52	96.3	10	14	US-10-351-641-1143	Sequence 1143, Ap
17	52	96.3	10	14	US-10-351-641-1309	Sequence 1309, Ap
18	52	96.3	10	14	US-10-351-641-1344	Sequence 1344, Ap
19	52	96.3	10	14	US-10-311-688-4	Sequence 4, Appli
20	52	96.3	10	15	US-10-170-096A-23	Sequence 23, Appli
21	52	96.3	10	15	US-10-360-101-1	Sequence 1, Appli
22	52	96.3	10	15	US-10-360-101-259	Sequence 259, App
23	52	96.3	10	15	US-10-353-160A-1	Sequence 1, Appli
24	52	96.3	10	15	US-10-617-561-1	Sequence 1, Appli
25	52	96.3	10	15	US-10-617-561-9	Sequence 9, Appli
26	52	96.3	10	16	US-10-298-378-1	Sequence 1, Appli
27	52	96.3	10	16	US-10-298-851-1	Sequence 1, Appli
28	52	96.3	11	13	US-10-044-034-17	Sequence 17, Appli
29	52	96.3	16	10	US-09-305-924-5	Sequence 5, Appli
30	52	96.3	17	10	US-09-305-924-3	Sequence 3, Appli
31	52	96.3	17	10	US-09-305-924-4	Sequence 4, Appli
32	52	96.3	18	14	US-10-351-641-1146	Sequence 1146, Ap
33	52	96.3	18	14	US-10-351-641-1147	Sequence 1147, Ap
34	52	96.3	18	14	US-10-351-641-1148	Sequence 1148, Ap
35	52	96.3	18	14	US-10-351-641-1172	Sequence 1172, Ap
36	52	96.3	18	14	US-10-351-641-1173	Sequence 1173, Ap
37	52	96.3	20	10	US-09-964-201A-26	Sequence 26, Appli
38	52	96.3	20	10	US-09-964-201A-29	Sequence 29, Appli
39	52	96.3	20	10	US-09-964-201A-30	Sequence 30, Appli
40	52	96.3	20	10	US-09-964-201A-31	Sequence 31, Appli
41	52	96.3	22	14	US-10-351-641-1145	Sequence 1145, Ap
42	52	96.3	26	14	US-10-351-641-1144	Sequence 1144, Ap
43	52	96.3	27	14	US-10-076-674-7	Sequence 7, Appli
44	52	96.3	27	15	US-10-355-161A-7	Sequence 7, Appli
45	52	96.3	28	9	US-09-848-834A-11	Sequence 11, Appli

ALIGNMENTS

RESULT 1
US-09-019-010-2
; Sequence 2, Application US/09019010
; Patent No. US20010014330A1
; GENERAL INFORMATION:
; APPLICANT: HARLAND, RICHARD
; APPLICANT: MANN, JOHN G.
; APPLICANT: ACRES, STEPHEN D.
; TITLE OF INVENTION: IMMUNIZATION AGAINST ENDOGENOUS MOLECULES
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: ROBINS & ASSOCIATES
; STREET: 90 MIDDLEFIELD ROAD, SUITE 200
; CITY: MENLO PARK
; STATE: CA
; COUNTRY: USA
; ZIP: 94025
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/019,010
; FILING DATE: 05-FEB-1998
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/036,883
; FILING DATE: 05-FEB-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: MCGACKEN, THOMAS P.
; REGISTRATION NUMBER: 38,548
; REFERENCE/DOCKET NUMBER: 9001-0035
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (650) 325-7812
; TELEFAX: (650) 325-7823
; INFORMATION FOR SEQ ID NO: 2:

FH Key Location/Qualifiers
 FT Misc-difference 1 /label= Pyr
 FT Modified-site 10
 FT /label= Gly-NH2
 XX BE887639-A.
 PN
 XX
 PD 24-AUG-1981.
 XX
 XX 22-MAY-1980; 80US-00152241.
 PF
 XX 22-MAY-1980; 80US-00152241.
 PR
 XX (AMHP) AYERST MCKENNA & HARRISON LTD.
 PA
 XX Auclair C;
 PI
 XX WPI; 1981-66067D/37.
 DR
 XX
 XX Gonadorelin for treatment of benign prostatic hyperplasia - is the
 PT deca:peptide Pyr-His-Trp-Ser-Tyr-Gly-Leu-Arg-Pro-Gly-NH2 or luteinising
 ET hormone liberating hormone.
 FT
 XX Claim 1; Page 7; 9pp; French.
 PS
 XX The inventors claim a compsn. for the redn. or prevention of undesired
 XX prostatic growth in males. The compsn. contains a decapeptide
 CC (gonadorelin) (API0097) with an appropriate vehicle or support. The
 CC compsn. is used for treating e.g. benign prostatic hyperplasia by
 CC parenteral admin. in daily doses of 0.035-11.0 (pref. 0.080-2.0) mg/kg.
 CC Gonadorelin is the generic name for LH-RH and is described in US3835108.
 CC In the example s.c. injection of the compsn. significantly reduced the wt.
 CC of seminal vesicles and ventral prostate in rats without affecting the
 CC wt. of the testicles. (Updated on 25-MAR-2003 to correct PA field.)
 CC (Updated on 27-AUG-2003 to correct OS field.)
 CC
 XX Sequence 10 AA;
 SQ

Query Match 96.3%; Score 52; DB 1; Length 10;
 Best Local Similarity 100.0%; Pred. No. 0.0078;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 HWSYGLRP 9
 |||||
 Db 2 HWSYGLRP 9

Search completed: March 10, 2004, 09:12:00
 Job time : 18.0584 secs

DT DT
XX XX
DE DE
DD DD

Luteinising Hormone Releasing Hormone analogue #5.

LHRH; Follicle Stimulating Factor; FSH; acne; hirsutism; dysmenorrhea;
precocious puberty; endometriosis; prostate cancer;
benign prostatic hypertrophy; mammary tumour.
Mammalia.

Synthetic.

Key Key Location/Qualifiers
Modified-site /label= OTHER
FT FT Modified-site /note= "pyroglutamic acid"
FT FT Modified-site /label= OTHER
FT FT Modified-site /note= "N-alpha-methyl-Leu"
FT FT Modified-site /note= "amidated or absent, in which case Pro(9) is Pro-NH-CZHS"

BE885308-A.
19-MAR-1981.
21-SEP-1979; 79FR-00023545.
21-SEP-1979; 79FR-00023545.
(ROUS) ROUSSEL-UCLAF.
Labrie P, Raynaud J;
WPI; 1981-23409D/14.
LH-RH, liberating factor for LH and FSH, and its agonists compen. - used
to treat prostate adenocarcinoma, benign hypertyrophy of the prostate,
hirsutism, acne, etc.
Claim 1(f); Page 16; 27pp; French.

A composition is claimed containing LHRH or its analogues. The
composition is used to treat prostate adenocarcinoma, benign hypertrophy
of the prostate, endometriosis, dysmenorthea, hirsutism, hormone-
dependent mammary tumours, for treatment or prevention of precocious
puberty, delaying the onset of puberty and for treating acne. The
compositions may also contain antiandrogens. See AAP10411-P10418.
(Updated on 01-JUL-2002 to add missing FI field.) (Updated on 10-MAR-2003
to add missing OS field.) (Updated on 25-MAR-2003 to correct PA field.)

Query Match 96.3%; Score 52; DB 1; Length 10;
Best Local Similarity 100.0%; Pred.No. 0.0078;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY QY 2 HWSYGRLP 9
DB DB 2 HWSYGRLP 9

RESULT 14
AAP10411 ID AAP10411 standard; peptide; 10 AA.

AC AC AAP10411;
XX XX 25-MAR-2003 (revised)
XX XX 10-MAR-2003 (revised)
XX XX 01-JUL-2002 (revised)
DT DT 17-DEC-1992 (first entry)

CC antihyperlipaemic; antithrombotic and/or anticancer agent or an agent for
 CC treating menopause. The composition of the invention is useful for
 CC treating and preventing diseases associated with angiotensin II, such as:
 CC circulatory diseases; hypertension; gastrointestinal disorders and
 CC cancer. The present amino acid sequence is used in the exemplification of
 CC the invention

XX SQ Sequence 9 AA;

Query Match 96.3%; Score 52; DB 6; Length 9;
 Best Local Similarity 100.0%; Pred. No. 1.4e+06;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 HWSYGLRP 9
 |||||
 DB 2 HWSYGLRP 9

RESULT 11
 AAE29840
 ID AAE29840 standard; peptide; 9 AA.

XX AC AAE29840;

XX DT 24-FEB-2003 (first entry)

XX DE Gonadotrophin releasing hormone analogue (GnRH-A) #3.

XX KW Gonadotrophin releasing hormone analogue; neurotoxin; prostate cancer;
 KW endocrine disorder; gonadotrophin related illness; endometrial cancer;
 KW pancreatic cancer; breast cancer; endometriosis; precocious puberty;
 KW GnRH-A; therapy.

XX OS Unidentified.

XX FH Key Location/Qualifiers

FT Modified-site 1 /note= "Pyroglutamic acid"

FT Modified-site 9 /note= "C-terminal ethylamide"

XX WO200274327-A2.

XX PD 26-SEP-2002.

XX PF 11-MAR-2002; 2002WO-US007379.

XX PR 15-MAR-2001; 2001US-00810601.

XX PA (ALLER) ALLERGAN SALES INC.

XX PI Donovan S;

XX DR WPI; 2003-018772/01.

XX PT New agent comprising a light chain and a (modified) heavy chain of a
 PT botulinum, butyricum, or tetani toxin, useful for treating a
 PT gonadotrophin related illness, e.g. breast, prostate pancreatic or
 PT endometrial cancer, or endometriosis.

XX PS Disclosure; Page 29; 97pp; English.

XX CC The invention relates to an agent comprising a neurotoxin preferably
 CC botulinum toxin for treating endocrine disorders for e.g. gonadotrophin
 CC related illness. The agent is useful for treating gonadotrophin related
 CC illness e.g. prostate cancer, endometrial cancer, pancreatic cancer,
 CC breast cancer, endometriosis or precocious puberty. It is also useful for
 CC decreasing gonadotrophin secretion in a mammal. The present sequence is
 CC gonadotrophin releasing hormone analogue (GnRH-A)

XX SQ Sequence 9 AA;

Query Match 96.3%; Score 52; DB 6; Length 9;

Best Local Similarity 100.0%; Pred. No. 1.4e+06;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 HWSYGLRP 9
 |||||
 DB 2 HWSYGLRP 9

RESULT 12

ABP96021
 ID ABP96021 standard; peptide; 9 AA.

XX AC ABP96021;

XX DT 02-MAY-2003 (first entry)

XX DE Gonadotrophin releasing hormone agonist peptide gonadrelin.

XX KW Gonadotrophin releasing hormone agonist; GnRH agonist; cancer;
 KW premenopausal breast cancer; cytostatic.

XX OS Synthetic.

XX FH Key Location/Qualifiers

FT Modified-site 9

FT /note= "Pro is C-terminally modified with -NHCH2CH3"

XX WO200287616-A1.

XX PD 07-NOV-2002.

XX PF 24-APR-2002; 2002WO-JP004071.

XX PR 25-APR-2001; 2001JP-00128032.

XX PA (TAKE) TAKEDA CHEM IND LTD.

XX PI Igari Y, Kusaka M;

XX DR WPI; 2003-148286/14.

XX PT Agent for preventing post-operative recurrence of premenopausal breast
 PT cancer contains GnRH agonists or antagonists.

XX PS Disclosure; Page 7; 39pp; Japanese.

XX CC The present invention describes an agent containing gonadotrophin
 CC releasing hormone (GnRH) agonists or antagonists. GnRH agonist and
 CC antagonists have cytostatic activity. The agent can be used for
 CC preventing post-operative recurrence of premenopausal breast cancer. The
 CC agents have no serious side effects with sustained drug effect over a
 CC long period without frequent administration. The present sequence
 CC represents an example of a GnRH agonist, from the present invention

XX SQ Sequence 9 AA;

Query Match 96.3%; Score 52; DB 6; Length 9;

Best Local Similarity 100.0%; Pred. No. 1.4e+06;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 HWSYGLRP 9
 |||||
 DB 2 HWSYGLRP 9

RESULT 13

AAP10416

ID AAP10416 standard; peptide; 10 AA.

XX AC AAP10416;

XX DT 25-MAR-2003 (revised)

DT 10-MAR-2003 (revised)

XX Bridon DP, Ezrin AM, Milner PG, Holmes DL, Thibaudau K;
 XX WPI; 2001-112059/12.
 XX
 XX Modifying and attaching therapeutic peptides to albumin prevents
 PT peptidase degradation, useful for increasing length of in vivo activity.
 XX
 XX Disclosure; Page 238; 733pp; English.
 XX
 XX The present invention describes a modified therapeutic peptide (I)
 CC comprising a therapeutically active amino acid region (III) and a
 CC reactive group (II) (e.g. succinimidyl and maleimido groups) attached to
 CC a less therapeutically active amino acid region (IV), which covalently
 CC bonds with amino/hydroxy/thiol groups on blood components to form a
 CC peptidase stabilised therapeutic peptide composed of 3-50 amino acids.
 CC (I) are useful for modifying therapeutic peptides e.g. hormones, growth
 CC factors and neurotransmitters, to protect them from peptidase activity in
 CC vivo for the treatment of various disorders. Endogenous therapeutic
 CC peptides are not suitable as drug candidates as they require frequent
 CC administration due to rapid degradation by peptidases in the body.
 CC Modifying and attaching therapeutic peptides to albumin prevents or
 CC reduces the action of peptidases to increase length of activity (half
 CC life) and specificity as bonding to large molecules decreases
 CC intracellular uptake and interference with physiological processes.
 CC AAB90829 to AAB92441 represent peptides which can be used in the
 CC exemplification of the present invention
 XX
 XX Sequence 9 AA;
 SQ

Query Match 96.3%; Score 52; DB 4; Length 9;
 Best Local Similarity 100.0%; Pred. No. 1.4e+06;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 HWSYGLRP 9
 |||||
 DB 2 HWSYGLRP 9

RESULT 9
 AAB59836
 ID AAB59836 standard; peptide; 9 AA.
 AC AAB59836;
 XX
 XX 26-MAR-2001 (first entry)
 DT
 DE GnRH peptide.
 XX
 XX GnRH-III; autoimmune disease; transplant rejection; retroviral disease;
 KW graft-versus-host-disease; lymphoproliferative disease;
 KW gonadotropin-releasing hormone.
 XX
 XX Petromyzon marinus.
 OS
 XX
 XX Key Location/Qualifiers
 FH Modified-site 1 /note= "Linked to Glucagon-like peptide"
 FT
 FT Modified-site 9
 FT Modified-site /note= "Modified with NH-CH3"
 XX
 XX WO200074724-A2.
 PN
 XX
 XX 14-DEC-2000.
 PD
 XX
 XX 05-JUN-2000; 2000WO-GB002014.
 PF
 XX
 XX 03-JUN-1999; 99GB-00012807.
 PR
 XX 03-JUN-1999; 99US-0137592P.
 XX
 XX (BIOI-) BIO INNOVATION LTD.
 PA
 XX Franks CR, Della Bitta R, Maitland NJ, Knight DJ;
 PI
 XX WPI; 2001-061658/07.
 DR

XX Novel product comprising proliferatively active moiety linked to genetic
 PT material, useful as vectors for protected nucleic acid material and as
 PT mitogen to stimulate proliferation of target cell.
 XX
 XX Disclosure; Page 4; 49pp; English.
 XX
 XX The present invention relates to a product comprising a proliferatively
 CC active moiety (PAM) linked to nucleic acid material which is associated
 CC with a protective material. The PAM product is useful for manufacturing a
 CC medicament for treating e.g. an autoimmune disease, transplant rejection,
 CC retroviral disease, graft-versus-host-disease, or lymphoproliferative
 CC disease, comprising cells bearing a high affinity receptor for PAM. The
 CC present sequence is a peptide of gonadotropin-releasing hormone (GnRH).
 CC GnRH is a peptide hormone, which has high-affinity receptors, and
 CC therefore can be used in the present invention
 XX
 XX Sequence 9 AA;
 SQ

Query Match 96.3%; Score 52; DB 4; Length 9;
 Best Local Similarity 100.0%; Pred. No. 1.4e+06;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 HWSYGLRP 9
 |||||
 DB 1 HWSYGLRP 8

RESULT 10
 AAO16785
 ID AAO16785 standard; peptide; 9 AA.
 XX
 AC AAO16785;
 XX
 XX 04-AUG-2003 (first entry)
 DT
 DE Sustained release composition-related peptide #5.
 XX
 XX Sustained release composition; angiotensin II antagonist; menopause;
 KW angiotensin II-related disease; circulatory disease; hypertension;
 KW gastrointestinal disorder; cancer.
 XX
 XX Unidentified.
 OS
 XX
 XX Key Location/Qualifiers
 FH Modified-site 1 /note= "Pyro glutamic acid"
 FT
 FT Modified-site 9
 FT Modified-site /note= "Modified with NH-CH3"
 XX
 XX WO2003013609-A1.
 PN
 XX
 XX 20-FEB-2003.
 PD
 XX
 XX 01-AUG-2002; 2002WO-JP007862.
 PF
 XX
 XX 03-AUG-2001; 2001JP-00236794.
 PR
 XX (TAKE) TAKEDA CHEM IND LTD.
 PA (KAWA/) KAWAMURA R.
 XX
 XX Kusumoto K, Hoshino T;
 PI
 XX
 XX WPI; 2003-256500/25.
 DR
 XX Sustained release composition comprises angiotensin II antagonist and
 PT antihypertensive, antihypoglycemic, antihyperlipemic, antithrombotic or
 PT anticancer agent or agent for treating menopause.
 XX
 XX Disclosure; Page 37; 120pp; Japanese.
 PS
 XX The invention comprises sustained release compositions that contain:
 CC angiotensin II antagonist; antihypertensive; antihypoglycaemic;
 CC

XX DE Amino acid sequence of truncated luteinising hormone releasing hormone.
 XX KW T helper cell epitope; CDV; immune response; canine vaccine;
 KW luteinising hormone releasing hormone; LHRH.
 XX OS Canis sp.
 XX PN WO200046390-A1.
 XX PD 10-AUG-2000.
 XX PF 07-FEB-2000; 2000WO-AU000070.
 XX PR 05-FEB-1999; 99AU-00008533.
 PR 04-AUG-1999; 99AU-00002013.
 XX PA (UYME) UNIV MELBOURNE.
 PA (CSLC-) CSL LTD.
 PA (CSIR) COMMONWEALTH SCI & IND RES ORG.
 PA (COUN-) COUNCIL QUEENSLAND INST MEDICAL RES.
 PA (HALL-) HALL INST MEDICAL RES WALTER & ELIZA.
 XX PI Jackson DC, Souravi G, Walker J;
 XX WPI; 2000-532904/48.
 XX Novel T helper cell epitopes derived from canine distemper virus useful
 PT for preparation of canine vaccines.
 XX Example 3; Page 21; 54pp; English.
 XX The present sequence represents luteinising hormone releasing hormone
 CC (LHRH). It is used in vaccines with T helper cell epitopes AAB08076-
 CC B08101, derived from canine distemper virus (CDV). Compositions
 CC comprising these T cell helper epitopes are useful for inducing an immune
 CC response in an animal. The epitopes are useful as components of animal,
 CC in particular, canine vaccines, either simply as synthetic peptide based
 CC vaccines and as additions to vaccines containing more complex antigens
 XX Sequence 9 AA;
 SQ
 Query Match 96.3%; Score 52; DB 3; Length 9;
 Best Local Similarity 100.0%; Pred. No. 1.4e+06;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 2 HWSYGLRP 9
 Db 1 HWSYGLRP 8
 RESULT 7
 AAB90979
 ID AAB90979 standard; peptide; 9 AA.
 XX AC AAB90979;
 XX 22-JUN-2001 (first entry)
 DE Luteinising hormone releasing hormone (LH-RH) related peptide SEQ ID:153.
 KW Protection; endogenous therapeutic peptide; peptidase; conjugation;
 KW blood component; modification; succinimidyl; maleimido group; amino;
 KW hydroxyl; thiol; hormone; growth factor; neurotransmitter.
 XX Homo sapiens.
 OS Synthetic.
 XX WO200069900-A2.
 XX 23-NOV-2000.
 XX 17-MAY-2000; 2000WO-US013576.
 PF

XX PR 17-MAY-1999; 99US-0134406P.
 PR 10-SEP-1999; 99US-0153406P.
 PR 15-OCT-1999; 99US-0159783P.
 XX (CONJ-) CONJUCHEM INC.
 XX Bridon DP, Ezrin AM, Milner PG, Holmes DL, Thibaudau K;
 PI WPI; 2001-112059/12.
 XX Modifying and attaching therapeutic peptides to albumin prevents
 PT peptidase degradation, useful for increasing length of in vivo activity.
 XX Disclosure; Page 240; 733pp; English.
 XX The present invention describes a modified therapeutic peptide (I)
 CC comprising a therapeutically active amino acid region (IIT) and a
 CC reactive group (II) (e.g. succinimidyl and maleimido groups) attached to
 CC a less therapeutically active amino acid region (IV), which covalently
 CC bonds with amino/hydroxyl/thiol groups on blood components to form a
 CC peptidase stabilised therapeutic peptide composed of 3-50 amino acids.
 CC (I) are useful for modifying therapeutic peptides e.g. hormones, growth
 CC factors and neurotransmitters, to protect them from peptidase activity in
 CC vivo for the treatment of various disorders. Endogenous therapeutic
 CC peptides are not suitable as drug candidates as they require frequent
 CC administration due to rapid degradation by peptidases in the body.
 CC Modifying and attaching therapeutic peptides to albumin prevents or
 CC reduces the action of peptidases to increase length of activity (half
 CC life) and specificity as bonding to large molecules decreases
 CC intracellular uptake and interference with physiological processes.
 CC AAB90829 to AAB92441 represent peptides which can be used in the
 CC exemplification of the present invention
 XX Sequence 9 AA;
 SQ
 Query Match 96.3%; Score 52; DB 4; Length 9;
 Best Local Similarity 100.0%; Pred. No. 1.4e+06;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 2 HWSYGLRP 9
 Db 1 HWSYGLRP 8
 RESULT 8
 AAB90972
 ID AAB90972 standard; peptide; 9 AA.
 XX AC AAB90972;
 XX 22-JUN-2001 (first entry)
 DE Luteinising hormone releasing hormone (LH-RH) related peptide SEQ ID:146.
 KW Protection; endogenous therapeutic peptide; peptidase; conjugation;
 KW blood component; modification; succinimidyl; maleimido group; amino;
 KW hydroxyl; thiol; hormone; growth factor; neurotransmitter.
 XX Homo sapiens.
 OS Synthetic.
 XX WO200069900-A2.
 XX 23-NOV-2000.
 XX 17-MAY-2000; 2000WO-US013576.
 XX 17-MAY-1999; 99US-0134406P.
 PR 10-SEP-1999; 99US-0153406P.
 PR 15-OCT-1999; 99US-0159783P.
 XX (CONJ-) CONJUCHEM INC.
 PA

Best Local Similarity 100.0%; Pred. No. 1.4e+06; Indels 0; Gaps 0;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 HWSYGLRP 9
| | | | |
DB 1 HWSYGLRP 8

RESULT 5
AAB15363
ID AAB15363 standard; peptide; 9 AA.
XX
AC AAB15363;
XX
DT 17-JAN-2001 (first entry)
XX
DE Human LHRH peptide SEQ ID NO: 2.
XX
KW Human; LHRH; GnRH; luteinising hormone releasing hormone;
gonadotrophin releasing hormone; fertility control; cancer;
endometriosis; prostate enlargement.
XX
OS Homo sapiens.
XX
PN WO200041720-A1.
XX
PD 20-JUL-2000.
XX
PF 24-DEC-1999; 99WO-AU001167.
XX
PR 08-JAN-1999; 99AU-00008073.
XX
PA (CSLC-) CSL LTD.
XX
PI Walker J;
XX
DR WPI; 2000-475954/41.
XX
PT Adjuvant composition for manufacturing an immunogenic composition that
can elicit an immune response in an animal, comprises an ionic
polysaccharide component and a saponin component that is an
immunostimulating complex.
XX
PS Disclosure; Page 50; 53pp; English.
XX
CC The present sequence is a peptide fragment of human luteinising hormone
releasing hormone (also known as LHRH, GnRH and gonadotrophin releasing
hormone). It was used to demonstrate the novel adjuvant of the invention,
which has lower reactivity than previous compositions. Vaccination of
humans and animals against LHRH can be used as a method of fertility
control, as well as enabling the control and treatment of disorders of
the reproductive organs, such as testicular, breast, prostate and ovarian
cancers, prostate enlargement and endometriosis. The composition of the
invention contains an anionic macromolecule and a saponin component, the
latter of which is an immunostimulant, and it can also be used with other
immunogens including soluble protein antigens, peptide haptens conjugated
to a carrier protein and whole viruses
XX
SQ Sequence 9 AA;
Query Match 96.3%; Score 52; DB 3; Length 9;
Best Local Similarity 100.0%; Pred. No. 1.4e+06;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 HWSYGLRP 9
| | | | |
DB 1 HWSYGLRP 8

RESULT 6
AAB08104
ID AAB08104 standard; peptide; 9 AA.
XX
AC AAB08104;
XX
DT 04-DEC-2000 (first entry)

Best Local Similarity 100.0%; Pred. No. 1.4e+06; Indels 0; Gaps 0;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 HWSYGLRP 9
| | | | |
DB 1 HWSYGLRP 8

RESULT 4
AAW94891
ID AAW94891 standard; peptide; 9 AA.
XX
AC AAW94891;
XX
DT 11-MAY-1999 (first entry)
XX
DE LHRH peptide fragment.
XX
KW LHRH; immune response; luteinising hormone releasing hormone; DT;
diphtheria toxin; castrating; oestrus cycling; aggression; breast;
sexual activity; organoleptic; livestock; cell growth; malignant;
prostate; ovarian; oncofetal; hyperplastic; pregnancy; endometriosis;
inflammatory response.
XX
OS Homo sapiens.
XX
PN WO9902180-A1.
XX
PD 21-JAN-1999.
XX
PF 09-JUL-1998; 98WO-AU000532.
XX
PR 09-JUL-1997; 97AU-00007768.
XX
PA (CSLC-) CSL LTD.
XX
PI Monanara MK;
XX
DR WPI; 1999-120511/10.
XX
PT New immunogenic luteinising hormone releasing hormone compositions -
comprise LHRH conjugated to diphtheria toxin and adsorbed to an ionic
polysaccharide, used to inhibit reproductive function in animals.
XX
PS Example 3; Page 30; 41pp; English.
XX
CC The invention relates immunogenic composition for eliciting an immune
response to luteinising hormone releasing hormone (LHRH). The composition
comprises a LHRH-diphtheria toxin (DT) conjugate adsorbed to an ionic
polysaccharide. The LHRH-DT compositions can be used for eliciting an
immune response to LHRH, for castrating an animal, for regulating oestrus
cycling in a female animal or for inhibiting characteristics induced by
the sexual maturation of an animal, e.g. aggression or sexual activity.
CC They can also be used for achieving production gains in livestock, e.g.
reduction or elimination of unwanted organoleptic characteristics from
the meat of livestock. They can also be used for inhibiting the growth of
cells which are regulated directly or indirectly by LHRH, e.g. malignant
breast cells, malignant prostate cells, malignant ovarian cells,
malignant oncofetal cells or hyperplastic cells. They can also be used
for down-regulating the libido of an animal. They can also be used for
inhibiting pregnancy, prostate enlargement, endometriosis or inflammatory
responses. The LHRH compositions induce a more effective immune response
against LHRH than the LHRH-carrier-adjuvant compositions. The effective
immune response against LHRH results in prevention of the release of the
hormones LH and FSH from the anterior pituitary. Sequences AAW94890-93
are peptide derivatives of LHRH
XX
SQ Sequence 9 AA;
Query Match 96.3%; Score 52; DB 2; Length 9;
Best Local Similarity 100.0%; Pred. No. 1.4e+06;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

CC activities. (I) can be used as GnRH agonists used in combination with
 CC selective oestrogen receptor modulators, selective androgen receptor
 CC modulators, sex hormone synthesis inhibitors, lyase inhibitors, receptor-
 CC type tyrosine kinase inhibitors, bone metabolism regulators,
 CC immunotherapy drugs, cytokine/chemokine inhibitors or endothelin receptor
 CC antagonists for treating and preventing endometriosis, myometrium tumour,
 CC Alzheimer's disease, circulatory system disorders, menopausal disorders,
 CC irregular periods, cancer metastasis, premenstrual syndrome, muscular
 CC distress, osteopathies due to calcium/phosphorus imbalance,
 CC prostatically or prostate cancer or breast cancer or their reoccurrence
 CC or metastasis. Combination enhances the quality of life by enhancing GnRH
 CC agonist activity and/or reducing side effects. The present sequence
 CC represents a GnRH agonist related peptide which is given in the present
 CC invention
 XX
 SQ Sequence 8 AA;

Query Match 96.3%; Score 52; DB 6; Length 8;
 Best Local Similarity 100.0%; Pred. No. 1.4e+06;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 HWSYGLRP 9
 Db 1 HWSYGLRP 8

RESULT 2

AAP10414
 ID AAP10414 standard; protein; 9 AA.

XX
 AC AAP10414;

DT 25-MAR-2003 (revised)
 DT 10-MAR-2003 (revised)
 DT 01-JUL-2002 (revised)
 DT 17-DEC-1992 (first entry)

DE Luteinising Hormone Releasing Hormone analogue #3.

XX [LHRH] Follicle Stimulating Factor; FSH; acne; hirsutism; dysmenorrhea;
 KW premature pubert; endometriosis; prostate cancer;
 KW benign prostate hypertrophy; mammary tumour.

XX Mammalia.

OS Mammalia.

XX Key Location/Qualifiers

FT Modified-site 1 /label= OTHER

FT Modified-site 9 /note= "pyroglutamic acid"

FT Modified-site 9 /note= "Pro-NH-(CH2)n-CH3 (n=0-2), Pro-NH-(CH2)2-OH or
 protected by pyrrolidino or morpholino gp."

XX Mammalia

XX 19-MAR-1981.

XX 79FR-00023545

XX 79FR-00023545.

XX 79FR-00023545.

XX 79FR-00023545.

XX 79FR-00023545.

XX 79FR-00023545.

XX 79FR-00023545.

XX 79FR-00023545.

XX 79FR-00023545.

XX 79FR-00023545.

XX 79FR-00023545.

XX 79FR-00023545.

XX A composition is claimed containing LHRH or its analogues. The
 CC composition is used to treat prostate adenocarcinoma, benign hypertrophy
 CC of the prostate, endometriosis, dysmenorrhea, hirsutism, hormone-
 CC dependent mammary tumours, for treatment or prevention of precocious
 CC puberty, delaying the onset of puberty and for treating acne. The
 CC compositions may also contain antiandrogens. See AAP10411-P10418.
 CC (Updated on 01-JUL-2002 to add missing PI field.) (Updated on 10-MAR-2003
 CC to add missing OS field.) (Updated on 25-MAR-2003 to correct PA field.)

XX Sequence 9 AA;

Query Match 96.3%; Score 52; DB 1; Length 9;

Best Local Similarity 100.0%; Pred. No. 1.4e+06;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 HWSYGLRP 9

Db 2 HWSYGLRP 9

RESULT 3

AAP50568

ID AAP50568 standard; protein; 9 AA.

XX
 AC AAP50568;

DT 03-OCT-2002 (revised)

DT 29-NOV-1991 (first entry)

DE Sequence of gonadoliberin analogue IIC.

XX Gonadoliberin agonist; follitropin release; lutropin release;

KW parathormone; hypertension; therapy.

XX Unidentified.

XX Key Location/Qualifiers

FT Modified-site 1 /label= pyroGlu

FT Modified-site 9 /label= bonded to -NHCH3, -NH-CH2-CH3, -NH-CH2CH2CH3

XX DE3332329-A.

XX 28-MAR-1985.

XX 08-SEP-1983; 83DE-03332329.

XX 08-SEP-1983; 83DE-03332329.

XX (FARH) HOECHST AG.

XX Konig W, Neubauer H;

XX WPI; 1985-081717/14.

XX Compsn. contg. Gonadoliberin or its analogues - for treating parathormone

XX deficiency states, e.g. hypocalcaemic conditions or hypertension.

XX Disclosure; Page 6-7; 17pp; German.

XX The inventors claim a composn. for treating metabolic disorders caused by

XX inadequate secretion of endogenous parathormone (PTH) which contains,

XX apart from an acceptable carrier, gonadoliberin or agonists at least as

XX strongly active as gonadoliberin. For parenteral use these provide 0.5-5

XX micrograms gonadoliberin per unit dose, and for application to mucosa

XX (intranasally) 10-200 micrograms per dose, for an adult of average wt.

XX (Updated on 03-OCT-2002 to add missing OS field.)

XX Sequence 9 AA;

XX Query Match 96.3%; Score 52; DB 1; Length 9;

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 10, 2004, 08:58:48 ; Search time 15.0584 Seconds
(without alignments)
187.635 Million cell updates/sec

Title: US-09-848-834A-1
Perfect score: 54
Sequence: 1 XHWSYGLRPX 10

Scoring table:
BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_29Jan04.*

1: Geneseq1980s.*
2: Geneseq1990s.*
3: Geneseq2000s.*
4: Geneseq2001s.*
5: Geneseq2002s.*
6: Geneseq2003as.*
7: Geneseq2003bs.*
8: Geneseq2004s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	52	96.3	8	6 ABP96550	Abp96550 Gonadotropin
2	52	96.3	9	1 AAP10414	Aap10414 Luteinisi
3	52	96.3	9	1 AAP50568	Aap50568 Sequence
4	52	96.3	9	2 AAW94891	Aaw94891 LHRH pept
5	52	96.3	9	3 AAB15363	Aab15363 Human LHR
6	52	96.3	9	3 AAB08104	Aab08104 Amino aci
7	52	96.3	9	4 AAB90979	Aab90979 Luteinisi
8	52	96.3	9	4 AAB90972	Aab90972 Luteinisi
9	52	96.3	9	4 AAB59836	Aab59836 GnRH pept
10	52	96.3	9	6 AAO16785	Aao16785 Sustained
11	52	96.3	9	6 AAE29840	Aae29840 Gonadotro
12	52	96.3	9	6 ABP96021	Abp96021 Gonadotro
13	52	96.3	10	1 AAP10416	Aap10416 Luteinisi
14	52	96.3	10	1 AAP10411	Aap10411 Luteinisi
15	52	96.3	10	1 AAP10097	Aap10097 Sequence
16	52	96.3	10	1 AAP20277	Aap20277 Modified
17	52	96.3	10	1 AAP50222	Aap50222 Gonadotro
18	52	96.3	10	1 AAP60127	Aap60127 Gonadolib
19	52	96.3	10	1 AAP61403	Aap61403 Gonadotro
20	52	96.3	10	1 AAP60576	Aap60576 Novel dec
21	52	96.3	10	1 AAP70922	Aap70922 Luteinisi
22	52	96.3	10	1 AAP90630	Aap90630 Sequence
23	52	96.3	10	2 AAR15713	Aar15713 Peptide #
24	52	96.3	10	2 AAR26819	Aar26819 LH releas
25	52	96.3	10	2 AAR33434	Aar33434 Therapeut

ALIGNMENTS

RESULT 1

ABP96550
ID ABP96550 standard; peptide; 8 AA.

XX AC ABP96550;

XX DT 27-MAY-2003 (first entry)

XX DE Gonadotropin releasing hormone agonist gonadrelin peptide.

XX KW Gonadotropin releasing hormone agonist; GnRH agonist; combination drug; pharmaceutical; breast cancer; endometriosis; myometrium tumour;
KW Alzheimer's disease; circulatory system disorder; menopausal disorder;
KW irregular period; cancer metastasis; premenstrual syndrome; osteopathy;
KW muscular distress; calcium/phosphorus imbalance; SEM; gynaecological;
KW selective oestrogen receptor modulator; cytostatic; nootropic; muscular;
KW neuroprotective; cardiovascular; endocrine; osteopathic; prostatically;
XX prostate cancer.

XX OS Synthetic.

XX PN WO2003015820-A1.

XX PD 27-FEB-2003.

XX PF 08-AUG-2002; 2002WO-JP008130.

XX PR 10-AUG-2001; 2001JP-00244616.

XX PA (TAKE) TAKEDA CHEM IND LTD.

XX PI Furuya S, Kusaka M;

XX DR WPI; 2003-300573/29.

XX PT Pharmaceutical composition e.g. for breast cancer comprises gonadotropin releasing hormone agonist and selective estrogen receptor modulator.

XX FS Disclosure; Page 9; 73pp; Japanese.

XX CC The present invention describes a pharmaceutical composition (I) for treating breast cancer, endometriosis, myometrium tumour, Alzheimer's disease, circulatory system disorders, menopausal disorders, irregular periods, cancer metastasis, premenstrual syndrome, muscular distress or osteopathies due to calcium/phosphorus imbalance. (I) comprises a gonadotropin releasing hormone (GnRH) agonist and a selective oestrogen receptor modulator (SERM). (I) has cytostatic, gynaecological, nootropic, neuroprotective, cardiovascular, endocrine, muscular and osteopathic

26	52	96.3	10	2	AAR62689	Aar62689 LHRH bapc
27	52	96.3	10	2	AAR91197	Aar91197 LHRH pept
28	52	96.3	10	2	AAR75152	Aar75152 Gonadotro
29	52	96.3	10	2	AAR66845	Aar66845 Gonadotro
30	52	96.3	10	2	ADC79469	Adc79469 Naturally
31	52	96.3	10	2	AAW65201	Aaw65201 Luteinisi
32	52	96.3	10	2	AAW65203	Aaw65203 Luteinisi
33	52	96.3	10	2	AAW45642	Aaw45642 Luteinisi
34	52	96.3	10	2	AAW16390	Aaw16390 Gonadotro
35	52	96.3	10	2	AAW22390	Aaw22390 Gonadotro
36	52	96.3	10	2	AAW04612	Aaw04612 Luteinisi
37	52	96.3	10	2	AAW76373	Aaw76373 Rat GnRH
38	52	96.3	10	2	AAW76381	Aaw76381 Rat modif
39	52	96.3	10	2	AAW79566	Aaw79566 GnRH-1 po
40	52	96.3	10	2	AAW61541	Aaw61541 Peptide h
41	52	96.3	10	2	AAW96765	Aaw96765 Luteinisi
42	52	96.3	10	2	AAV03856	Aav03856 Amino aci
43	52	96.3	10	2	AAV03864	Aav03864 Amino aci
44	52	96.3	10	2	AAV31180	Aav31180 Ubiquitin
45	52	96.3	10	2	AAV31176	Aav31176 Ubiquitin

Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 HWSYGLRP 9
| | | | |
Db 66 HWSYALSP 73

RESULT 15

Q86D88 PRELIMINARY; PRT; 165 AA.
AC Q86D88;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Preprogonadotropin-releasing hormone 2 precursor.
OS Ciona intestinalis.
OC Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea; Enterogona;
OC Phlebobranchia; Clonidae; Ciona.
OX NCBI_TaxID=7719;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=22583575; PubMed=12697698;
RA Adams B.A., Tello J.A., Erchegyi J., Warby C., Hong D.J.,
RA Akinsanya K.O., Mackie G.O., Vale W., Rivier J.E., Sherwood N.M.;
RT "Six Novel Gonadotropin-Releasing Hormones Are Encoded as Triplets on
RT Each of Two Genes in the Protochordate, Ciona intestinalis.";
RL Endocrinology 144:1907-1919(2003).
DR EMBL; AY204708; AAP06795.1; -;
DR GO; GO:0005576; C:extracellular; IEA.
DR GO; GO:0005179; P:hormone activity; IEA.
DR GO; GO:0007275; P:development; IEA.
DR InterPro; IPR002012; GNRH.
DR Pfam; PF00446; GNRH; 1.
DR PROSITE; PS00473; GNRH; 3.
KW Signal.
FT SIGNAL 1 63 Potential.
FT CHAIN 66 75 gonadotropin-releasing hormone 7.
FT CHAIN 79 88 gonadotropin-releasing hormone 8.
FT CHAIN 92 101 gonadotropin-releasing hormone 4.
FT CHAIN 107 165 GnRH-associated peptide.
SQ SEQUENCE 165 AA; 19250 MW; CE3F748849BE3601 CRC64;

Query Match 74.1%; Score 40; DB 5; Length 165;

Best Local Similarity 75.0%; Pred. No. 14;

Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 HWSYGLRP 9
| | | | |
Db 67 HWSYALSP 74

Search completed: March 10, 2004, 09:25:19

Job time : 11.8833 secs

```

RL Submitted (JUN-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY320285; AAP84607.1; -.
FT NON_TER 1
SQ SEQUENCE 72 AA; 8064 MW; 0FAAC44E2AB792BF CRC64;

Query Match 77.8%; Score 42; DB 13; Length 72;
Best Local Similarity 75.0%; Pred. No. 2.3;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 HWSYGLRP 9
DB 2 HWSFGLSP 9

RESULT 12
ID Q8JH60 PRELIMINARY; PRT; 86 AA.
AC Q8JH60;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Gonadotropin-releasing hormone.
OS Alosa sapidissima (American shad).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Clupeomorpha; Clupeidae;
OC Alosa.
OX NCBI_TaxID=34773;
RN [1]
RP SEQUENCE FROM N.A.
RA Abraham E., Gochilf Y., Zohar Y.;
RT "American shad (Alosa sapidissima) hrGnRH sequence.";
RL Submitted (AUG-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF363681; AAN04492.1; -.
DR GO; GO:0005576; C:extracellular; IEA.
DR GO; GO:0005179; F:hormone activity; IEA.
DR GO; GO:0007275; P:development; IEA.
DR InterPro; IPR002012; GnRH.
DR Pfam; PF00446; GnRH.1.
DR PROSITE; PS00473; GnRH; 1.
SQ SEQUENCE 86 AA; 9554 MW; 8E4921F3CF2350E3 CRC64;

Query Match 75.9%; Score 41; DB 13; Length 86;
Best Local Similarity 75.0%; Pred. No. 4.3;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 HWSYGLRP 9
DB 24 HWSHGLSP 31

RESULT 13
ID Q8G815 PRELIMINARY; PRT; 1127 AA.
AC Q8G815;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Carbamoyl-phosphate synthase large chain.
CN CARB OR BL0068.
OS Bifidobacterium longum.
OC Bacteria; Actinobacteria; Actinobacteridae; Bifidobacteriales;
OC Bifidobacteriaceae; Bifidobacterium.
OX NCBI_TaxID=216816;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=NCC 2705.
RX MEDLINE=22294977; PubMed=12381787;
RA Schell M.A., Kamrantsou M., Snel B., Vilanova D., Berger B.,
RA Pessi G., Zwaalen M.-C., Desiere F., Bork P., Delley M.,
RA Pridmore R.D., Arigoni F.;
RT "The genome sequence of Bifidobacterium longum reflects its adaptation
to the human gastrointestinal tract.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:14422-14427(2002).
```

```

DR EMBL; AE014621; AAN23934.1; -.
DR GO; GO:0005737; C:cytoplasm; IEA.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0004056; F:carbamoyl-phosphate synthase activity; IEA.
DR GO; GO:0016874; F:ligase activity; IEA.
DR GO; GO:0006526; F:arginine biosynthesis; IEA.
DR GO; GO:0006807; F:nitrogen metabolism; IEA.
DR GO; GO:0019856; F:pyrimidine base biosynthesis; IEA.
DR InterPro; IPR006275; CarA_L_glu.
DR InterPro; IPR005483; CPase_L.
DR InterPro; IPR005479; CPase_L_D2.
DR InterPro; IPR005480; CPase_L_D3.
DR InterPro; IPR005481; CPase_L_N.
DR InterPro; IPR004362; MGS_like.
DR Pfam; PF00289; CPase_L_Chain; 2.
DR Pfam; PF02786; CPase_L_D2; 3.
DR Pfam; PF02787; CPase_L_D3; 1.
DR Pfam; PF02142; MGS_1.
DR PRINTS; PR00098; CPASE.
DR TIGRFAMs; TIGR01369; CPaseL1_lrg; 1.
DR PROSITE; PS00866; CPASE_1; 2.
DR PROSITE; PS00867; CPASE_2; 2.
KW Complete proteome.
SQ SEQUENCE 1127 AA; 123194 MW; 2B3C0B4FE3282489 CRC64;

Query Match 75.9%; Score 41; DB 16; Length 1127;
Best Local Similarity 85.7%; Pred. No. 74;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 3 WSYGLRP 9
DB 518 WTYGLRP 524

RESULT 14
Q86D87 PRELIMINARY; PRT; 164 AA.
AC Q86D87;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Preprogonadotropin-releasing hormone 2 precursor (fragment).
OS Ciona intestinalis.
OC Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea; Enterogona;
OC Phlebobranchia; Cionidae; Ciona.
OX NCBI_TaxID=7719;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=22583575; PubMed=12697698;
RA Adams B.A., Tello J.A., Erchegyi J., Warby C., Hong D.J.,
RA Akinsanya K.O., Mackie G.O., Vale W., Rivier J.E., Sherwood N.M.;
RT "Six Novel Gonadotropin-Releasing Hormones Are Encoded as Triplets on
Each of Two Genes in the Protochordate, Ciona intestinalis.";
RL Endocrinology 144:1307-1319(2003).
DR EMBL; AY204709; AAP06796.1; -.
DR GO; GO:0005576; C:extracellular; IEA.
DR GO; GO:0005179; F:hormone activity; IEA.
DR GO; GO:0007275; P:development; IEA.
DR InterPro; IPR002012; GnRH.
DR Pfam; PF00446; GnRH; 1.
DR PROSITE; PS00473; GnRH; 3.
KW Signal.
RN NON_TER 1
FT SIGNAL <1 62 Potential.
FT CHAIN 65 74 gonadotropin-releasing hormone 7.
FT CHAIN 78 87 gonadotropin-releasing hormone 8.
FT CHAIN 91 100 gonadotropin-releasing hormone 4.
FT CHAIN 106 133 GnRH-associated peptide.
FT CHAIN 134 158 GnRH-associated peptide.
SQ SEQUENCE 164 AA; 19119 MW; C2962E474f4163C1 CRC64;

Query Match 74.1%; Score 40; DB 5; Length 164;
Best Local Similarity 75.0%; Pred. No. 13;
```

DB 25 HWSYGLSP 32

RESULT 10
Q92YR6 PRELIMINARY; PRT; 374 AA.

ID Q92YR6
AC Q92YR6;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Putative muconate cycloisomerase (EC 5.5.1.1).
GN RA0797 OR SMA1461.
OS Rhizobium meliloti (Sinorhizobium meliloti).
OC Plasmid pS_{MA} (megaplasmid).
CG Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
CC Rhizobiaceae; Sinorhizobium/Ensifer group; Sinorhizobium.
OX NCBI_TaxID=382;
OX [1]
RN SEQUENCE FROM N.A.
RP STRAIN=1021.
RC MEDLINE=21396509; PubMed=11481432;
RX Barnett M.J., Fisher R.F., Jones T., Komp C., Abola A.P.,
RA Barclay-Hubler F., Bowser L., Capela D., Galibert F., Gouzy J.,
RA Gurjal M., Hong A., Huizar L., Hyman R.W., Kahn D., Kahn M.L.,
RA Kalman S., Keating D.H., Palm C., Peck M.C., Surzycki R., Wells D.H.,
RA Yeh K.-C., Davis R.W., Federspiel N.A., Long S.R.;
RT "Nucleotide sequence and predicted functions of the entire
Sinorhizobium meliloti pS_{MA} megaplasmid";
RL Proc. Natl. Acad. Sci. U.S.A. 98:9883-9888(2001).
DR ENBL; AB007266; AAK65455.1; --
DR FIR; E95361; E95361.
DR GO; GO:0046821; C:extrachromosomal DNA; IEA.
DR GO; GO:0016853; F:isomerase activity; IEA.
DR GO; GO:0018849; F:muconate cycloisomerase activity; IEA.
DR GO; GO:0008152; P:metabolism; IEA.
DR InterPro: IPR001354; MR_MLE.
DR Pfam: PF01188; MR_MLE; 1.
DR Pfam: PF02746; MR_MLE.N; 1.
DR PROSITE; PS00509; MR_MLE.2; 1.
DR Isomerase; Plasmid; Hypothetical protein; Complete proteome.
KW SEQUENCE 374 AA; 40999 MW; 0369AA67790BE23D6 CRC64;
SQ

Query Match 79.6%; Score 43; DB 16; Length 374;
Best Local Similarity 85.7%; Pred. NO. 9.4;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0

Cy 2 HWSYGLSP 8
DB 21 HWSYGLSP 27

RESULT 11
Q7TIL2 PRELIMINARY; PRT; 72 AA.

ID Q7TIL2
AC Q7TIL2;
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Gonadotropin-releasing hormone (fragment).
OS Odonesthes bonariensis.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
CC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Acanthopterygii; Percomorpha; Atherinomorpha;
CC Atheriniformes; Atherinidae; Atherinidae; Atherinopsinae;
OC Odonesthes.
OX NCBI_TaxID=219752;
OX [1]
RN SEQUENCE FROM N.A.
RP TISSUE=Brain;
RC
FA Guilgur L.G., Miranda L.A., Somoza G.M.;
RT "Characterization of three GnRH cDNA sequences in the pejerrey fish
RT Odonesthes bonariensis.";

RD ADMITTED 2 MAR 82 08 00 AMM/ GEN/ 0500
 DR EMBL; AF467291; AM90320.1; -
 DR EMBL; AB104861; BAC65154.1; -

```

DR PRINTS; PR01541; GONADOLIBRN1.
DR PROSITE; PS00473; GNRH; 1.
KW Amidation; Hormone; Signal.
FT SIGNAL 1 22 POTENTIAL.
FT CHAIN 23 32 MGNRH.
FT CHAIN 33 91 GNRH ASSOCIATED PEPTIDE.
SQ SEQUENCE 91 AA; 9893 MW; BA15C9DC08434A7B CRC64;

Query Match 96.3%; Score 52; DB 13; Length 91;
Best Local Similarity 100.0%; Pred. No. 0.043;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 HWSYGLRP 9
Db 24 HWSYGLRP 31

RESULT 2
Q8JIF3 PRELIMINARY; PRT; 64 AA.
AC Q8JIF3; 22, Created
DT 01-OCT-2002 (TRENBLrel. 22, Last sequence update)
DT 01-OCT-2002 (TRENBLrel. 22, Last sequence update)
DT 01-JUN-2003 (TRENBLrel. 24, Last annotation update)
DE Gonadotropin-releasing hormone (Fragment).
OS Dentex dentex.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Acanthopterygii; Percormorpha; Perciformes; Percoidae;
OC Sparidae; Dentex.
OX NCBI_TaxID=94951;

SEQUENCE FROM N.A.
Kato M., Elmesiry G.E.;
"Sequence comparison of GnRH genes in closely-related Sparidae
fishes.";
Submitted (AUG-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB089313; BAC07229.1; -.
DR GO; GO:0005576; C:extracellular; IEA.
DR GO; GO:0005183; F:luteinizing hormone-releasing factor activity; IEA.
DR GO; GO:0007275; P:development; IEA.
DR InterPro; IPR002012; GNRH.
DR InterPro; IPR004079; Gonadolibrini.
DR Pfam; PF00446; GNRH; 1.
DR PRINTS; PR01541; GONADOLIBRN1.
DR PROSITE; PS00473; GNRH; 1.
FT NON_TER 1 64
SQ SEQUENCE 64 AA; 7028 MW; 1F91FFFB2B4BB6D0 CRC64;

Query Match 85.2%; Score 46; DB 13; Length 64;
Best Local Similarity 87.5%; Pred. No. 0.37;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 HWSYGLRP 9
Db 16 HWSYGLRP 23

RESULT 3
Q8JIF2 PRELIMINARY; PRT; 64 AA.
AC Q8JIF2; 22, Created
DT 01-OCT-2002 (TRENBLrel. 22, Last sequence update)
DT 01-OCT-2002 (TRENBLrel. 22, Last sequence update)
DT 01-JUN-2003 (TRENBLrel. 24, Last annotation update)
DE Gonadotropin-releasing hormone (Fragment).
OS Pagrus major (Red sea bream) (Chrysophrys major).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Acanthopterygii; Percormorpha; Perciformes; Percoidae;
OC Sparidae; Pagrus.
OX NCBI_TaxID=143350;

```

```

[1]
RN SEQUENCE FROM N.A.
RP Kato M., Elmesiry G.E.;
RT "Sequence comparison of GnRH genes in closely-related Sparidae
fishes.";
RL Submitted (AUG-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB089314; BAC07230.1; -.
DR GO; GO:0005576; C:extracellular; IEA.
DR GO; GO:0005183; F:luteinizing hormone-releasing factor activity; IEA.
DR GO; GO:0007275; P:development; IEA.
DR InterPro; IPR002012; GNRH.
DR InterPro; IPR004079; Gonadolibrini.
DR Pfam; PF00446; GNRH; 1.
DR PRINTS; PR01541; GONADOLIBRN1.
DR PROSITE; PS00473; GNRH; 1.
FT NON_TER 1 64
SQ SEQUENCE 64 AA; 6967 MW; 119F0D0CD14BB6D0 CRC64;

Query Match 85.2%; Score 46; DB 13; Length 64;
Best Local Similarity 87.5%; Pred. No. 0.37;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 HWSYGLRP 9
Db 16 HWSYGLRP 23

RESULT 4
Q8JIF4 PRELIMINARY; PRT; 68 AA.
AC Q8JIF4; 22, Created
DT 01-OCT-2002 (TRENBLrel. 22, Last sequence update)
DT 01-OCT-2002 (TRENBLrel. 22, Last sequence update)
DT 01-JUN-2003 (TRENBLrel. 24, Last annotation update)
DE Gonadotropin-releasing hormone (Fragment).
OS Acanthopagrus latus (Yellowfin porgy).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Acanthopterygii; Percormorpha; Perciformes; Percoidae;
OC Sparidae; Acanthopagrus.
OX NCBI_TaxID=8177;

SEQUENCE FROM N.A.
Kato M., Elmesiry G.E.;
"Sequence comparison of GnRH genes in closely-related Sparidae
fishes.";
RL Submitted (AUG-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB089312; BAC07228.1; -.
DR GO; GO:0005576; C:extracellular; IEA.
DR GO; GO:0005183; F:luteinizing hormone-releasing factor activity; IEA.
DR GO; GO:0007275; P:development; IEA.
DR InterPro; IPR002012; GNRH.
DR InterPro; IPR004079; Gonadolibrini.
DR Pfam; PF00446; GNRH; 1.
DR PRINTS; PR01541; GONADOLIBRN1.
DR PROSITE; PS00473; GNRH; 1.
FT NON_TER 1 68
SQ SEQUENCE 68 AA; 7543 MW; 067708609FB8E771 CRC64;

Query Match 85.2%; Score 46; DB 13; Length 68;
Best Local Similarity 87.5%; Pred. No. 0.4;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 HWSYGLRP 9
Db 20 HWSYGLRP 27

RESULT 5
Q9YI26 PRELIMINARY; PRT; 87 AA.
ID Q9YI26

```


GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 10, 2004, 08:58:54 ; Search time 9.88327 Seconds
(without alignments)
319.245 Million cell updates/sec

Title: US-09-848-834A-1
Perfect score: 54
Sequence: 1 XHWSVGLRPX 10

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues
Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

SPTREMBL_25:*

- 1: sp_archaea:*
- 2: sp_bacteria:*
- 3: sp_fungi:*
- 4: sp_human:*
- 5: sp_invertebrate:*
- 6: sp_mammal:*
- 7: sp_mhc:*
- 8: sp_organelle:*
- 9: sp_phage:*
- 10: sp_plant:*
- 11: sp_rodent:*
- 12: sp_virus:*
- 13: sp_vertebrate:*
- 14: sp_unclassified:*
- 15: sp_rvirus:*
- 16: sp_bacteriap:*
- 17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	52	96.3	91	13 Q9PRH0	Q9prh0 anguilla ja
2	46	85.2	64	13 Q8JIF3	Q8jif3 dentex dent
3	46	85.2	64	13 Q8JIF2	Q8jif2 pagrus major
4	46	85.2	68	13 Q8JIF4	Q8jif4 acanthopag
5	46	85.2	87	13 Q9YI26	Q9yi26 sparus aur
6	46	85.2	94	13 Q8JFY3	Q8jfy3 oreochromis
7	46	85.2	96	13 Q8UW80	Q8uw80 verasper mo
8	46	85.2	98	13 Q8U5A5	Q8u5a5 oreochromis
9	46	85.2	120	13 Q7T059	Q7t059 microgogoni
10	43	79.6	374	16 Q92YR6	Q92yr6 rhizobium m
11	42	77.8	72	13 Q7TIL2	Q7til2 odontesthes
12	41	75.9	86	13 Q8JH60	Q8jh60 alosa sapid
13	41	75.9	1127	16 Q8G815	Q8g815 bifidobacte
14	40	74.1	164	5 Q86DS7	Q86ds7 ciona intes
15	40	74.1	165	5 Q86DS8	Q86ds8 ciona intes
16	40	74.1	367	16 Q987W1	Q987w1 rhizobium l

17	40	74.1	1242	16 Q7UGU6	Q7ugu6 synechococc
18	40	74.1	1259	16 Q7V7L4	Q7v7l4 prochloroc
19	39	72.2	33	13 Q9W7G0	Q9w7g0 oncorhynch
20	39	72.2	33	13 Q9PT34	Q9pt34 oncorhynch
21	39	72.2	54	13 Q9W0W9	Q9w0w9 oncorhynch
22	39	72.2	62	13 Q90ZE1	Q90ze1 oncorhynch
23	39	72.2	68	13 Q7TIL1	Q7til1 odontesthes
24	39	72.2	82	13 Q92094	Q92094 oncorhynch
25	39	72.2	82	13 Q90VY3	Q90vy3 oncorhynch
26	39	72.2	82	13 Q9W7G1	Q9w7g1 oncorhynch
27	39	72.2	82	13 Q918P9	Q918p9 oncorhynch
28	39	72.2	82	13 Q918Q0	Q918q0 oncorhynch
29	39	72.2	88	13 Q9PSY9	Q9psy9 sparus aur
30	39	72.2	90	13 Q8UW82	Q8uw82 verasper mo
31	39	72.2	90	13 Q8AWF6	Q8awf6 oreochromis
32	39	72.2	90	13 Q7ZT00	Q7zt00 oreochromis
33	39	72.2	94	13 Q9DDD8	Q9ddd8 brachydanio
34	39	72.2	94	13 Q9DEH5	Q9deh5 carassius a
35	39	72.2	94	13 Q8UUK6	Q8uuk6 scleropages
36	39	72.2	94	13 Q9DEH6	Q9deh6 carassius a
37	39	72.2	94	13 Q8JHC3	Q8jhc3 cyprinus ca
38	39	72.2	94	13 Q804C1	Q804c1 cyprinus ca
39	39	72.2	94	13 Q801D6	Q801d6 cyprinus ca
40	39	72.2	94	13 Q801D5	Q801d5 cyprinus ca
41	39	72.2	388	17 Q9YD14	Q9ydl4 aeropyrum p
42	39	72.2	1461	5 Q8MYA8	Q8mya8 caenorhabdi
43	38	70.4	111	5 Q9VW18	Q9vw18 drosophila
44	38	70.4	161	10 Q9ZUG1	Q9zug1 arabidopsis
45	38	70.4	165	4 Q8NF82	Q8nf82 homo sapien

ALIGNMENTS

RESULT 1

Q9PRH0 PRELIMINARY; PRT; 91 AA.

AC Q9PRH0; 01-MAY-2000 (TREMREL. 13, Created)

DT 01-MAY-2000 (TREMREL. 13, Last sequence update)

DT 01-JUN-2003 (TREMREL. 24, Last annotation update)

DE Prepro-mGnRH precursor (Gonadoliberein) (Gonadotropin-releasing hormone) (LH-RH) (Luliberin)

OS Anguilla japonica (Japanese eel).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Anguilliformes; Anguillidae; Anguilla.

OX NCBI_TaxID=7937;

RN [1]_SEQUENCE FROM N.A.

RP SEQUENCE FROM N.A.

RC TISSUE=Brain;

RA Okubo K., Suetake H., Aida K.;

RT "Expression of two gonadotropin-releasing hormone (GnRH) precursor genes in various tissues of the Japanese eel and evolution of GnRH.";

RL Zool. Sci. 16:471-478(1999).

RN [2]_SEQUENCE FROM N.A.

RP SEQUENCE FROM N.A.

RA Okubo K., Suetake H., Aida K.;

RT "A splicing variant for the prepro-mammalian gonadotropin-releasing hormone (prepro-mGnRH) mRNA is present in the brain and various peripheral tissues of the Japanese eel.";

RL Zool. Sci. 16:645-651(1999).

CC -!- FUNCTION: STIMULATES THE SECRETION OF GONADOTROPINS (BY SIMILARITY).

CC -!- SIMILARITY: BELONGS TO THE GnRH FAMILY.

DR EMBL; AB028989; BAA8268.1; -.

DR EMBL; AB026991; BAA8397.1; -.

DR GO; GO:0005576; C:extracellular; IEA.

DR GO; GO:0005183; F:luteinizing hormone-releasing factor activity; IEA.

DR GO; GO:0007275; P:development; IEA.

DR InterPro; IPR002012; GnRH.

DR InterPro; IPR004079; GonadolibereinI.

DR Pfam; PF00446; GnRH; 1.

FEATURE:
OTHER INFORMATION: synthetic
NAME/KEY: misc feature
LOCATION: (1)..(1)
OTHER INFORMATION: Glu is modified with a pyro group.
NAME/KEY: misc feature
LOCATION: (10)..(10)
OTHER INFORMATION: Gly is modified with -NH2 group.
US-10-184-126-1

Query Match 96.3%; Score 52; DB 13; Length 10;
Best Local Similarity 100.0%; Pred. No. 0.022;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 HWSYGLRP 9
Db 2 HWSYGLRP 9

RESULT 13

US-10-115-553-1
Sequence 1, Application US/10115553
Publication No. US20030040482A1
GENERAL INFORMATION:

APPLICANT: Roeseke, Roger W.
TITLE OF INVENTION: LHRH Antagonist Peptides
FILE REFERENCE: PFI-007CPUS
CURRENT APPLICATION NUMBER: US/10/115,553
CURRENT FILING DATE: 2002-04-02
PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 08/973,378
PRIOR FILING DATE: EARLIER FILING DATE: 1998-04-06
PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 08/480,494
PRIOR FILING DATE: EARLIER FILING DATE: 1995-06-07
NUMBER OF SEQ ID NOS: 1
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 1
LENGTH: 10
TYPE: PRT
ORGANISM: Homo sapiens
US-10-115-553-1

Query Match 96.3%; Score 52; DB 14; Length 10;
Best Local Similarity 100.0%; Pred. No. 0.022;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 HWSYGLRP 9
Db 2 HWSYGLRP 9

RESULT 14

US-10-122-483-1
Sequence 1, Application US/10122483
Publication No. US20030044936A1
GENERAL INFORMATION:

APPLICANT: Hwang, Jaulang
APPLICANT: Hsu, Chia-Tse
APPLICANT: Ting, Chun-Jen
TITLE OF INVENTION: PEPTIDE REPEAT IMMUNOGENS
FILE REFERENCE: 08919-071001
CURRENT APPLICATION NUMBER: US/10/122,483
CURRENT FILING DATE: 2002-07-15
PRIOR APPLICATION NUMBER: 09/412,558
PRIOR FILING DATE: 1999-10-05
NUMBER OF SEQ ID NOS: 13
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 1
LENGTH: 10
TYPE: PRT
ORGANISM: Homo sapiens
US-10-122-483-1

Query Match 96.3%; Score 52; DB 14; Length 10;
Best Local Similarity 100.0%; Pred. No. 0.022;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 HWSYGLRP 9
Db 2 HWSYGLRP 9

RESULT 15

US-10-117-364-1
Sequence 1, Application US/10117364
Publication No. US20030181385A1
GENERAL INFORMATION:

APPLICANT: Roeseke, Roger W.
TITLE OF INVENTION: LHRH Antagonist Peptides
FILE REFERENCE: PFI-007CPUS
CURRENT APPLICATION NUMBER: US/10/117,364
CURRENT FILING DATE: 2002-04-05
PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: US/08/973,378
PRIOR FILING DATE: EARLIER FILING DATE: 1998-04-06
PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 08/480,494
PRIOR FILING DATE: EARLIER FILING DATE: 1995-06-07
NUMBER OF SEQ ID NOS: 1
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 1
LENGTH: 10
TYPE: PRT
ORGANISM: Homo sapiens
US-10-117-364-1

Query Match 96.3%; Score 52; DB 14; Length 10;
Best Local Similarity 100.0%; Pred. No. 0.022;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 HWSYGLRP 9
Db 2 HWSYGLRP 9

Search completed: March 10, 2004, 10:25:47
Job time : 8.85992 secs

```

OX NCBI_TaxID=143350;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RA Okuzawa K., Granneman J., Bogerd J., Goos H., Zohar Y., Kagawa H.;
RL Submitted (SEP-1996) to the EMBL/GenBank/DDAJ databases.
CC -!- FUNCTION: Stimulates the secretion of gonadotropins (By
CC similarity).
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- SIMILARITY: Belongs to the GnRH family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; D86582; BAA13129.1; -
DR InterPro; IPR002012; GnRH
DR InterPro; IPR004079; GonadoliberinI.
DR Pfam; PF00446; GnRH; 1.
DR PRINTS; PRO1541; GONADOLIBRN1.
DR PROSITE; PS00473; GnRH; 1.
KW Cleavage on pair of basic residues; Hormone; Amidation; Hypothalamus;
KW Signal; Multigene family; Pyroliidone carboxylic acid.
FT SIGNAL 1 23 POTENTIAL
FT CHAIN 24 95 PROGONADOLIBERIN I.
FT PEPTIDE 24 33 GONADOLIBERIN I.
FT PEPTIDE 37 95 GnRH-ASSOCIATED PEPTIDE I (POTENTIAL).
FT MOD_RES 24 24 PYROLIDONE CARBOXYLIC ACID (BY
FT MOD_RES 33 33 SIMILARITY).
FT MOD_RES 33 33 AMIDATION (G-34 PROVIDE AMIDE GROUP
FT MOD_RES 33 33 (BY SIMILARITY).
FT SEQUENCE 95 AA; 10566 MW; 61E79C990328D73E CRC64;
Query Match 85.2%; Score 46; DB 1; Length 95;
Best Local Similarity 87.5%; Pred. No. 0.069;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 2 HWSYGLRP 9
Db 25 HWSYGLSP 32
Search completed: March 10, 2004, 09:13:45
Job time : 2.82879 secs

```

RN SEQUENCE FROM N.A.
 RP MEDLINE=99061842; PubMed=9843638;
 RA White R.B., Fernald R.D.;
 RX "Ontogeny of gonadotropin-releasing hormone (GnRH) gene expression
 RT reveals a distinct origin for GnRH-containing neurons in the
 RL midbrain.";
 RL Gen. Comp. Endocrinol. 112:322-329 (1998).
 RN [3]
 RN SEQUENCE OF 23-32, AND MASS SPECTROMETRY.
 RC TISSUE=PIUITARY;
 RX MEDLINE=95372591; PubMed=7644702;
 RA Powell J.F.F., Fischer W.H., Park M., Craig A.G., Rivier J.E.,
 RX White S.A., Francis R.C., Fernald R.D., Licht P., Warby C.,
 RA Sherwood N.M.;
 RT "Primary structure of solitary form of gonadotropin-releasing hormone
 RT (GnRH) in chick pituitary; three forms of GnRH in brain of chick
 RT and pumpkinseed fish.";
 RL Regul. Pept. 57:43-53 (1995).
 CC -1- FUNCTION: STIMULATES THE SECRETION OF GONADOTROPINS. MAY BE
 CC RESPONSIBLE FOR THE REGULATION OF THE HYPOTHALAMIC-PIUITARY-
 CC GONADAL AXIS.
 CC -1- SUBCELLULAR LOCATION: Secreted.
 CC -1- TISSUE SPECIFICITY: SYNTHESIZED IN PREOPTIC NEURONS AND IS
 CC TRANSPORTED TO THE PIUITARY IN THE PREOPTIC-HYPOTHYSEAL AXONS.
 CC -1- MASS SPECTROMETRY: MW=1113.9; METHOD=MALDI; RANGE=23-32.
 CC -1- SIMILARITY: Belongs to the GnRH family.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; U31865; AAC59691.1; -;
 CC EMBL; AF076961; AAC27716.1; -;
 CC PIR; I50739; I50739.
 CC GO; GO:0005576; C:extracellular; NAS.
 CC GO; GO:0005183; F:luteinizing hormone-releasing factor activity; NAS.
 CC GO; GO:0007275; P:development; IDA.
 CC InterPro; IPR002012; GnRH.
 CC InterPro; IPR004079; GonadolibereinI.
 CC Pfam; PF00446; GnRH; 1.
 CC PRINTS; PR01541; GONADOLIBERNI.
 CC PROSITE; PS00473; GnRH; 1.
 CC Cleavage on pair of basic residues; Hormone; Amidation; Hypothalamus;
 CC Signal; Multigene family; Pyrrolidone carboxylic acid.
 CC SIGNAL 1 22 PROGNADOLIBERIN I.
 CC CHAIN 23 94 GONADOLIBERIN I.
 CC PEPTIDE 36 94 GnRH-ASSOCIATED PEPTIDE I (POTENTIAL).
 CC FT PEPTIDE 23 23 PYRROLIDONE CARBOXYLIC ACID.
 CC FT MOD_RES 32 32 AMIDATION (G-33 PROVIDE AMIDE GROUP).
 CC FT CONFLICT 86 94 ENGHRTPKK -> KMDTGHGRNERFL (IN REF. 1).
 CC SEQUENCE 94 AA; 10382 MW; E57DBA83333278D7 CRC64;
 CC
 CC Query Match 85.2%; Score 46; DB 1; Length 94;
 CC Best Local Similarity 87.5%; Pred. No. 0.068;
 CC Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 CC
 CC QY 2 HWSYGLRP 9
 CC |||||
 CC 24 HWSYGLSP 31
 CC
 CC RESULT 14
 CC GONL MORSA
 CC ID GONL MORSA STANDARD; PRT; 95 AA.
 CC AC Q73812.
 CC DT 16-OCT-2001 (Rel. 40, Created)
 CC DT 16-OCT-2001 (Rel. 40, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Gonadoliberein I precursor (Gonadotropin-releasing hormone I) (GnRH-I)
 DE (LH-RH I) (Luliberin I).
 GN GNRH1.
 OS Morone saxatilis (Striped bass).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
 OC Acanthomorpha; Acanthopterygii; Percomorpha; Perciformes; Percoidae;
 OC Moronidae; Morone.
 OC NCBI_TaxID=34816;
 RN [1]
 RN SEQUENCE FROM N.A.
 RP MEDLINE=99061809; PubMed=9845669;
 RX Chow M.M., Kight K.E., Gotthelf Y., Alok D., Stubblefield J., Zohar Y.;
 RA "Multiple GnRHs present in a teleost species are encoded by separate
 RT genes: analysis of the sbGnRH and cGnRH-II genes from the striped
 RT bass, Morone saxatilis.";
 RL J. Mol. Endocrinol. 21:277-289 (1998).
 CC -1- FUNCTION: Stimulates the secretion of gonadotropins (By
 CC similarity).
 CC -1- SUBCELLULAR LOCATION: Secreted.
 CC -1- SIMILARITY: Belongs to the GnRH family.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; AF056314; AAD03817.1; -;
 CC InterPro; IPR002012; GnRH.
 CC InterPro; IPR004079; GonadolibereinI.
 CC Pfam; PF00446; GnRH; 1.
 CC PRINTS; PR01541; GONADOLIBERNI.
 CC PROSITE; PS00473; GnRH; 1.
 CC Cleavage on pair of basic residues; Hormone; Amidation; Hypothalamus;
 CC Signal; Multigene family; Pyrrolidone carboxylic acid.
 CC SIGNAL 1 22 POTENTIAL.
 CC CHAIN 23 95 PROGNADOLIBERIN I.
 CC PEPTIDE 23 32 GONADOLIBERIN I.
 CC FT PEPTIDE 36 95 GnRH-ASSOCIATED PEPTIDE I (POTENTIAL).
 CC FT PEPTIDE 23 23 PYRROLIDONE CARBOXYLIC ACID (BY
 CC SIMILARITY).
 CC FT MOD_RES 32 32 AMIDATION (G-33 PROVIDE AMIDE GROUP)
 CC FT MOD_RES 32 32 (BY SIMILARITY).
 CC SEQUENCE 95 AA; 10411 MW; 980C6988FC279BFC CRC64;
 CC
 CC Query Match 85.2%; Score 46; DB 1; Length 95;
 CC Best Local Similarity 87.5%; Pred. No. 0.069;
 CC Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 CC
 CC QY 2 HWSYGLRP 9
 CC |||||
 CC 24 HWSYGLSP 31
 CC
 CC RESULT 15
 CC GONL PAGMA
 CC ID GONL PAGMA STANDARD; PRT; 95 AA.
 CC AC P70074;
 CC DT 15-JUL-1998 (Rel. 36, Created)
 CC DT 15-JUL-1998 (Rel. 36, Last sequence update)
 CC DT 28-FEB-2003 (Rel. 41, Last annotation update)
 CC DE Gonadoliberein I precursor (Gonadotropin-releasing hormone I) (GnRH-I)
 CC (LH-RH I) (Luliberin I).
 GN GNRH1.
 OS Pagrus major (Red sea bream) (Chrysophrys major).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
 OC Acanthomorpha; Acanthopterygii; Percomorpha; Perciformes; Percoidae;
 OC Sparidae; Pagrus.

DE Gonadoliberin I (Gonadotropin-releasing hormone I) (GnRH-I) (LH-RH I)
 DS (Luliberin I).
 OS Alligator mississippiensis (American alligator).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Archosauria; Crocodylidae; Alligatorinae; Alligator.
 OX NCBI_TaxID=8496;
 RN [1]
 RP SEQUENCE.
 RC TISSUE=Brain;
 RX MEDLINE=91352338; PubMed=1882082;
 RA Lovejoy D.A., Fischer W.H., Parker D.B., McRory J.E., Park M.,
 RA Lance V., Swanson P., Rivier J.E., Sherwood N.M.;
 RT "Primary structure of two forms of gonadotropin-releasing hormone
 from brains of the American alligator (Alligator mississippiensis).";
 RL Regul. Pept. 33:105-116(1991).
 CC -!- FUNCTION: Stimulates the secretion of gonadotropins.
 CC -!- SUBCELLULAR LOCATION: Secreted.
 CC -!- SIMILARITY: Belongs to the GnRH family.
 DR PIR: A60066; RHAQL.
 DR InterPro: IPR002012; GnRH.
 DR Pfam: PF00446; GnRH; 1.
 DR PROSITE: PS00473; GnRH; 1.
 KW Hormone; Amidation; Hypothalamus; Pyrrolidone carboxylic acid.
 FT MOD_RES 1 1 PYRROLIDONE CARBOXYLIC ACID.
 FT MOD_RES 10 10 AMIDATION.
 SQ SEQUENCE 10 AA; 1172 MW; 284823D7286B45A3 CRC64;
 Query Match 88.9%; Score 48; DB 1; Length 10;
 Best Local Similarity 87.5%; Pred. No. 0.003;
 Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
 Qy 2 HWSYGLRP 9
 Db 2 HWSYGLQP 9
 RESULT 12
 GONI_CHICK
 ID GONI_CHICK STANDARD; PRT; 92 AA.
 AC P37042; P20407;
 DT 01-FEB-1991 (Rel. 17, Created)
 DT 01-JUN-1994 (Rel. 29, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Gonadoliberin I precursor [Contains: Gonadoliberin I (LH-RH I)
 DE (Luteinizing hormone-releasing hormone I) (Gonadotropin-releasing
 hormone I) (GnRH I) (Luliberin I); GnRH-associated peptide II].
 OS Gallus gallus (Chicken).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
 OC Gallus.
 OX NCBI_TaxID=9031;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=White leghorn;
 RX MEDLINE=94059355; PubMed=7902095;
 RA Dunn I.C., Chen Y., Hook C., Sharp P.J., Sang H.M.;
 RT "Characterization of the chicken pregonadotropin-releasing
 hormone-1 gene.";
 RL J. Mol. Endocrinol. 11:19-29(1993).
 RN [2]
 RP SEQUENCE OF 24-33.
 RC TISSUE=Hypothalamus;
 RX MEDLINE=82265778; PubMed=7050119;
 RA King J.A., Millar R.P.;
 RT "Structure of chicken hypothalamic luteinizing hormone-releasing
 hormone. II. Isolation and characterization.";
 RL J. Biol. Chem. 257:10729-10732(1982).
 RN [3]
 RP SEQUENCE OF 24-33.
 RC TISSUE=Hypothalamus;
 RX MEDLINE=82265777; PubMed=7050118;
 RA King J.A., Millar R.P.;
 RT "Structure of chicken hypothalamic luteinizing hormone-releasing
 hormone. I. Structural determination on partially purified
 material.";
 RL J. Biol. Chem. 257:10722-10728(1982).
 CC -!- FUNCTION: Stimulates the secretion of gonadotropins.
 CC -!- SUBCELLULAR LOCATION: Secreted.
 CC -!- SIMILARITY: Belongs to the GnRH family.
 CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation.
 CC the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.ebi-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 DR EMBL: X69491; CAA49246.1; -.
 DR PIR: I50644; I50644.
 DR GO: GO:0005576; C:extracellular; IDA.
 DR GO: GO:0005183; F:luteinizing hormone-releasing factor activity; IDA.
 DR GO: GO:0007275; F:development; IDA.
 DR InterPro: IPR002012; GnRH.
 DR Pfam: PF00446; GnRH; 1.
 DR PRINTS: PR01541; GONADOLIBERNI.
 DR PROSITE: PS00473; GnRH; 1.
 KW Cleavage on pair of basic residues; Hormone; Amidation; Hypothalamus;
 KW Signal; Pyrrolidone carboxylic acid.
 FT SIGNAL 1 23
 FT CHAIN 24 92 PROGNADOLIBERIN I.
 FT PEPTIDE 24 33 GONADOLIBERIN I.
 FT PEPTIDE 37 92 GnRH-ASSOCIATED PEPTIDE I.
 FT MOD_RES 24 24 PYRROLIDONE CARBOXYLIC ACID.
 FT MOD_RES 33 33 AMIDATION (G-34 PROVIDE AMIDE GROUP).
 SQ SEQUENCE 92 AA; 10206 MW; 61AE7EBAF508B6A CRC64;
 Query Match 88.9%; Score 48; DB 1; Length 92;
 Best Local Similarity 87.5%; Pred. No. 0.028;
 Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
 Qy 2 HWSYGLRP 9
 Db 25 HWSYGLQP 32
 RESULT 13
 GONI_HAPBU
 ID GONI_HAPBU STANDARD; PRT; 94 AA.
 AC P51918; O93387;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Progonadoliberin I precursor [Contains: Gonadoliberin I (Luteinizing
 hormone releasing hormone I) (Gonadotropin-releasing hormone I)
 (GnRH-I) (LH-RH I) (Luliberin I); GnRH-associated peptide I].
 OS Haplochromis burtoni (Burton's mouthbrooder).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Necteleostei;
 OC Acanthomorpha; Acanthopterygii; Perciformes; Labroidae;
 OC Cichlidae; Astatotilapia.
 OX NCBI_TaxID=8153;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=95396797; PubMed=7667296;
 RA White S.A., Kasten T.L., Bond C.T., Adelman J.P., Fernald R.D.;
 RT "Three gonadotropin-releasing hormone genes in one organism suggest
 novel roles for an ancient peptide";
 RL Proc. Natl. Acad. Sci. U.S.A. 92:8363-8367(1995).

RN [4]
 RP SYNTHESIS OF 24-33.
 RX MEDLINE=82265777; PubMed=7050118;
 RA King J.A., Millar R.P.;
 RT "Structure of chicken hypothalamic luteinizing hormone-releasing
 hormone. I. Structural determination on partially purified
 material.";
 RL J. Biol. Chem. 257:10722-10728(1982).
 CC -!- FUNCTION: Stimulates the secretion of gonadotropins.
 CC -!- SUBCELLULAR LOCATION: Secreted.
 CC -!- SIMILARITY: Belongs to the GnRH family.
 CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation.
 CC the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.ebi-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 DR EMBL: X69491; CAA49246.1; -.
 DR PIR: I50644; I50644.
 DR GO: GO:0005576; C:extracellular; IDA.
 DR GO: GO:0005183; F:luteinizing hormone-releasing factor activity; IDA.
 DR GO: GO:0007275; F:development; IDA.
 DR InterPro: IPR002012; GnRH.
 DR Pfam: PF00446; GnRH; 1.
 DR PRINTS: PR01541; GONADOLIBERNI.
 DR PROSITE: PS00473; GnRH; 1.
 KW Cleavage on pair of basic residues; Hormone; Amidation; Hypothalamus;
 KW Signal; Pyrrolidone carboxylic acid.
 FT SIGNAL 1 23
 FT CHAIN 24 92 PROGNADOLIBERIN I.
 FT PEPTIDE 24 33 GONADOLIBERIN I.
 FT PEPTIDE 37 92 GnRH-ASSOCIATED PEPTIDE I.
 FT MOD_RES 24 24 PYRROLIDONE CARBOXYLIC ACID.
 FT MOD_RES 33 33 AMIDATION (G-34 PROVIDE AMIDE GROUP).
 SQ SEQUENCE 92 AA; 10206 MW; 61AE7EBAF508B6A CRC64;
 Query Match 88.9%; Score 48; DB 1; Length 92;
 Best Local Similarity 87.5%; Pred. No. 0.028;
 Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
 Qy 2 HWSYGLRP 9
 Db 25 HWSYGLQP 32
 RESULT 13
 GONI_HAPBU
 ID GONI_HAPBU STANDARD; PRT; 94 AA.
 AC P51918; O93387;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Progonadoliberin I precursor [Contains: Gonadoliberin I (Luteinizing
 hormone releasing hormone I) (Gonadotropin-releasing hormone I)
 (GnRH-I) (LH-RH I) (Luliberin I); GnRH-associated peptide I].
 OS Haplochromis burtoni (Burton's mouthbrooder).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Necteleostei;
 OC Acanthomorpha; Acanthopterygii; Perciformes; Labroidae;
 OC Cichlidae; Astatotilapia.
 OX NCBI_TaxID=8153;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=95396797; PubMed=7667296;
 RA White S.A., Kasten T.L., Bond C.T., Adelman J.P., Fernald R.D.;
 RT "Three gonadotropin-releasing hormone genes in one organism suggest
 novel roles for an ancient peptide";
 RL Proc. Natl. Acad. Sci. U.S.A. 92:8363-8367(1995).

```

RN  SEQUENCE FROM N.A.
RP  MEDLINE=89384661; PubMed=2476669;
RA  Bond C.T., Haylick J.S., Seeburg P.H., Adelman J.P.;
RT  "The rat gonadotropin-releasing hormone: SH locus: structure and
RT  hypothalamic expression.";
RL  Mol. Endocrinol. 3:1257-1262(1989).
RN  SEQUENCE FROM N.A.
RP  MEDLINE=93105480; PubMed=1468115;
RA  Mair C.C., Marchetti B., Leboeuf R.D., Bialock J.E.;
RT  "Mammalian genes transcribed from opposite strands of the same
RT  DNA locus.";
RL  Cell. Mol. Neurobiol. 12:447-454(1992).
RN  SEQUENCE OP 1-47 FROM N.A.
RP  MEDLINE=87149087; PubMed=3547652;
RA  Adelman J.P., Bond C.T., Douglass J., Herbert E.;
RT  "Two mammalian genes transcribed from opposite strands of the same
RT  DNA locus.";
RL  Science 235:1514-1517(1987).
CC  -!- FUNCTION: Stimulates the secretion of gonadotropins; it stimulates
CC  the secretion of both luteinizing and follicle-stimulating
CC  hormones.
CC  -!- SUBCELLULAR LOCATION: Secreted.
CC  -!- TISSUE SPECIFICITY: Central nervous system.
CC  -!- SIMILARITY: Belongs to the GnRH family.
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
DR  EMBL; G50870; AAB24572.1; -.
DR  EMBL; M12579; AAA41263.1; -.
DR  EMBL; M31670; AAA41264.1; -.
DR  EMBL; M15527; AAA42141.1; ALT_SEQ.
DR  EMBL; M15529; AAA42139.1; -.
DR  EMBL; M15528; -. NOT ANNOTATED_CDS.
DR  PIR; A40147; RHRTG.
DR  InterPro; IPR002012; GnRH.
DR  Pfam; PF00446; GnRH; 1.
DR  PRINTS; PR01541; GONADOLIBERN1.
DR  PROSITE; PS00473; GnRH; 1.
KW  Cleavage on pair of basic residues; Hormone; Amidation; Hypothalamus;
KW  Placenta; Signal; Pyrrolidone carboxylic acid.
FT  SIGNAL 1 23
FT  CHAIN 24 92
FT  PEPTIDE 24 33
FT  ACT_SITE 26 26
FT  MOD_RES 24 24
FT  MOD_RES 33 33
FT  SEQUENCE 92 AA; 10500 MW; 494B5C64DA8A3EB3 CRC64;
Query Match 96.3%; Score 52; DB 1; Length 92;
Best Local Similarity 100.0%; Pred. No. 0.005;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 2 HWSYGLRP 9
DB 25 HWSYGLRP 32
RESULT 10
GONI_TUPGB

```

```

ID  GONI_TUPGB STANDARD; PRT; 92 AA.
AC  Q95335;
DT  15-DEC-1998 (Rel. 37, Last sequence update)
DT  15-DEC-1998 (Rel. 37, Last sequence update)
DE  28-FEB-2003 (Rel. 41, Last annotation update)
DE  Gonadoliberin I precursor (Contains: Gonadoliberin I (LH-RH I)
DE  (Luteinizing hormone-releasing hormone I) (Gonadotropin-releasing
DE  hormone I) (GnRH I) (Luliberin I); GnRH-associated peptide I).
GN  GnRH1 OR GnRH.
OS  Tupia glis belangeri (Common tree shrew).
OC  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC  Mammalia; Eutheria; Scandentia; Tupaiidae; Tupia.
OX  NCBI_TaxID=37347;
RN  SEQUENCE FROM N.A.
RP  TISSUE=Hypothalamus;
RX  MEDLINE=87079639; PubMed=8921350;
RA  Kastan T.L., White S.A., Norton T.T., Bond C.T., Adelman J.P.;
RT  "Characterization of two new preproGnRH mRNAs in the tree shrew:
RT  first direct evidence for mesencephalic GnRH gene expression in a
RT  placental mammal.";
RL  Gen. Comp. Endocrinol. 104:7-19(1996).
CC  -!- FUNCTION: Stimulates the secretion of gonadotropins; it stimulates
CC  the secretion of both luteinizing and follicle-stimulating
CC  hormones.
CC  -!- SUBCELLULAR LOCATION: Secreted.
CC  -!- SIMILARITY: Belongs to the GnRH family.
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
DR  EMBL; U63326; AAB16837.1; -.
DR  InterPro; IPR002012; GnRH.
DR  Pfam; PF00446; GnRH; 1.
DR  PRINTS; PR01541; GONADOLIBERN1.
DR  PROSITE; PS00473; GnRH; 1.
KW  Cleavage on pair of basic residues; Hormone; Amidation; Hypothalamus;
KW  Placenta; Signal; Pyrrolidone carboxylic acid.
FT  SIGNAL 1 23
FT  CHAIN 24 92
FT  PEPTIDE 24 33
FT  PEPTIDE 37 92
FT  ACT_SITE 26 26
FT  MOD_RES 24 24
FT  MOD_RES 33 33
FT  SEQUENCE 92 AA; 10197 MW; 4FDBF2C58CF5F63B CRC64;
Query Match 96.3%; Score 52; DB 1; Length 92;
Best Local Similarity 100.0%; Pred. No. 0.005;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 2 HWSYGLRP 9
DB 25 HWSYGLRP 32
RESULT 11
GONI_ALLMI
ID  GONI_ALLMI STANDARD; PRT; 10 AA.
AC  P37041; P20407;
DT  01-FEB-1991 (Rel. 17, Created)
DT  01-FEB-1991 (Rel. 17, Last sequence update)
DT  28-FEB-2003 (Rel. 41, Last annotation update)

```

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).

CC EMBL; X01059; CAA35526.1; --
 CC EMBL; M12578; AAA35916.1; --
 CC EMBL; X15215; CAA33285.1; --
 CC PIR; S05308; RHHUG.
 CC Genew; HGNC:4419; GNRH1.
 CC MIM; 152760; --
 CC GO; GO:0005625; C:soluble fraction; TAS.
 CC GO; GO:0003883; P:luteinizing hormone-releasing factor activity; TAS.
 CC GO; GO:0007267; P:cell-cell signaling; TAS.
 CC GO; GO:0007275; P:development; TAS.
 CC GO; GO:0008285; P:negative regulation of cell proliferation; TAS.
 CC GO; GO:0007165; P:signal transduction; TAS.
 CC InterPro; IPR002012; GNRH.
 CC InterPro; IPR004079; Gonadoliberein1.
 CC Pfam; PF00446; GNRH_1.
 CC PRINTS; PRO1541; GONADOLIBERIN.
 CC PROSITE; PS00473; GNRH_1.
 CC Cleavage on pair of basic residues; Hormone; Amidation; Hypothalamus;
 CC Placenta; Pharmaceutical; Signal; Polymorphism;
 CC Pyroglutamate carboxylic acid.
 CC SIGNAL 1 23
 CC CHAIN 24 92
 CC MOD_RES 24 24
 CC MOD_RES 33 33
 CC PEPTIDE 24 33
 CC PEPTIDE 37 92
 CC ACT_SITE 26 26
 CC PROGNADOLIBERIN I.
 CC GONADOLIBERIN I.
 CC GNRH-ASSOCIATED PEPTIDE I.
 CC APPEARS TO BE ESSENTIAL FOR BIOLOGICAL
 CC ACTIVITY.
 CC PYROGLUTAMATE CARBOXYLIC ACID.
 CC AMIDATION (G-34 PROVIDE AMIDE GROUP).
 CC W -> S (in dSNP:6185).
 CC /FTid=VAR_013943.
 CC FT SEQUENCE 92 AA; 10380 MW; 30A72221E076FA79 CRC64;
 SQ
 Query Match 96.3%; Score 52; DB 1; Length 92;
 Best Local Similarity 100.0%; Pred. No. 0.005;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 2 HWSYGLRP 9
 Db 25 HWSYGLRP 32
 RESULT 9
 GON1_RAT
 ID GON1_RAT STANDARD; PRT; 92 AA.
 AC P07490;
 DT 01-APR-1988 (Rel. 07, Created)
 DT 01-APR-1988 (Rel. 07, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Progonadoliberein I precursor [Contains: Gonadoliberein I (LH-RH I)
 DE (luteinizing hormone-releasing hormone I) (Gonadotropin-releasing
 DE hormone I) (GNRH I) (Luliberin I) (Gonadorelin); GNRH-associated
 DE peptide I].
 DE GNRH1 OR GNRH OR LHRH.
 GN Homo sapiens (Human).
 OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OC NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RP MEDLINE=89366682; PubMed=2671939;
 RA Hayflick J.S., Adelman J.P., Seeburg P.H.;
 RT "The complete nucleotide sequence of the human gonadotropin-releasing
 RT hormone gene."
 RL Nucleic Acids Res. 17:6403-6403 (1989).
 RN [2]
 RP SEQUENCE FROM N.A.
 RP MEDLINE=86094338; PubMed=2867549;
 RA Adelman J.P., Mason A.J., Hayflick J.S., Seeburg P.H.;
 RT "Isolation of the gene and hypothalamic cDNA for the common precursor
 RT of gonadotropin-releasing hormone and prolactin release-inhibiting
 RT factor in human and rat."
 RL Proc. Natl. Acad. Sci. U.S.A. 83:179-183 (1986).
 RN [3]
 RP SEQUENCE FROM N.A., AND VARIANT SER-16.
 RP MEDLINE=85012739; PubMed=6090551;
 RA Seeburg P.H., Adelman J.P.;
 RT "Characterization of cDNA for precursor of human luteinizing hormone
 RT releasing hormone."
 RL Nature 311:666-668 (1984).
 RN [4]
 RP SEQUENCE OF 24-33.
 RP MEDLINE=83126573; PubMed=6760865;
 RA Tan L., Rouseau P.;
 RT "The chemical identity of the immunoreactive LHRH-like peptide
 RT biosynthesized in the human placenta."
 RL Biochem. Biophys. Res. Commun. 109:1061-1071 (1982).
 RN [5]
 RP VARIANT SER-16.
 RP MEDLINE=99318093; PubMed=10391209;
 RA Cargill M., Altschuler D., Ireland J., Sklar P., Ardlie K., Patil N.,
 RA Shaw N., Lane C.R., Lim E.P., Kalyanaram N., Nimesh J., Ziaugra L.,
 RA Friedland L., Rolfe A., Warrington J., Lipshutz R., Daley G.Q.,
 RA Lander E.S.;
 RT "Characterization of single-nucleotide polymorphisms in coding regions
 RT of human genes."
 RL Nat. Genet. 22:231-238 (1999).
 RN [6]
 RP ERRATUM.
 RA Cargill M., Altschuler D., Ireland J., Sklar P., Ardlie K., Patil N.,
 RA Shaw N., Lane C.R., Lim E.P., Kalyanaram N., Nimesh J., Ziaugra L.,
 RA Friedland L., Rolfe A., Warrington J., Lipshutz R., Daley G.Q.,
 RA Lander E.S.;
 RL Nat. Genet. 23:373-373 (1999).
 CC -!- FUNCTION: Stimulates the secretion of gonadotropins; it stimulates
 CC the secretion of both luteinizing and follicle-stimulating
 CC hormones.
 CC -!- SUBCELLULAR LOCATION: Secreted.
 CC -!- PHARMACEUTICAL: Available under the names Pactrel (Ayerst Labs),
 CC Lutreupse or Lutrelaf (Ferring Pharmaceuticals) and Relisorm
 CC (Serono).
 CC -!- SIMILARITY: Belongs to the GNRH family.

10-OCT-2003 (Rel. 42, Last annotation update)
 DE Progonadoliberin I precursor [Contains: Gonadoliberin I (LHRH I)
 DE (luteinizing hormone-releasing hormone I) (Gonadotropin-releasing
 DE (luteinizing hormone releasing hormone I) (Gonadotropin-releasing
 DE hormone I) (GnRH I) (Luliberin I); GnRH-associated peptide I (GAP1)].
 GN GNRH1 OR GNRH.
 OS Sus scrofa (Pig).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Amphibia; Batrachia; Anura; Neobatrachia; Ranidae; Rana.
 OC NCBI_TaxID=8400;
 RN NCBI_TaxID=8400;
 RN SEQUENCE FROM N.A., TISSUE SPECIFICITY, AND DEVELOPMENTAL STAGE.
 RC TISSUE=Forebrain; PubMed=11170016;
 RX MEDLINE=21102951; Wang L., Yoo M.S., Kang H.M., Im W.B., Choi H.S., Bogerd J.,
 RA "Cloning and characterization of cDNAs encoding the GnRH1 and GnRH2
 RT precursors from bullfrog (Rana catesbeiana).";
 RL J. Exp. Zool. 289:190-201(2001).
 CC -1- FUNCTION: Stimulates the secretion of gonadotropins (By
 CC similarity).
 CC -1- SUBCELLULAR LOCATION: Secreted.
 CC -1- TISSUE SPECIFICITY: Forebrain.
 CC -1- DEVELOPMENTAL STAGE: Expressed at significantly higher levels
 CC during post-breding. Not expressed in pituitary.
 CC -1- SIMILARITY: Belongs to the GnRH family.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; AF188754; AA05972.1; -;
 CC GO: GO:000576; C:extracellular; NAS.
 CC GO: GO:0005183; P:luteinizing hormone-releasing factor activity; NAS.
 CC GO: GO:0009755; P:hormone mediated signaling; NAS.
 CC GO: GO:0000003; P:reproduction; NAS.
 CC InterPro: IPR002012; GnRH.
 CC Pfam: PF00445; GnRH; 1.
 CC PRINTS: PR01541; GONADOLIBERNI.
 CC PROSITE; PS00473; GnRH; 1.
 CC Cleavage on pair of basic residues; Hormone; Amidation; Signal;
 CC Pyroglutamate carboxylic acid.
 FT SIGNAL 1 24
 FT CHAIN 25 90
 FT PEPTIDE 25 34
 FT PEPTIDE 38 86
 FT MOD_RES 25 25
 FT MOD_RES 34 34
 FT MOD_RES 34 34
 FT SEQUENCE 90 AA; 10291 MW; 317203B4E3DA3FE7 CRC64;
 Query Match 96.3%; Score 52; DB 1; Length 90;
 Best Local Similarity 100.0%; Pred. No. 0.0049;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 2 HWSYGLRP 9
 DB 26 HWSYGLRP 33
 RESULT 7
 ID GON1_PIG STANDARD; PRT; 91 AA.
 AC P49921;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)

Progonadoliberin I precursor [Contains: Gonadoliberin I (LHRH I)
 (luteinizing hormone-releasing hormone I) (Gonadotropin-releasing
 hormone I) (GnRH I) (Luliberin I); GnRH-associated peptide I].
 GN GNRH1 OR GNRH.
 OS Sus scrofa (Pig).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
 NCBI_TaxID=9823;
 RN NCBI_TaxID=9823;
 RN SEQUENCE FROM N.A.
 RC TISSUE=Hypothalamus;
 RA Weesner G.D., Mattern R.L., Becker B.A.;
 RL Submitted (MAY-1994) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RN SEQUENCE OF 24-33.
 RX MEDLINE=72114303; PubMed=4946067;
 RA Baba Y., Matsuo H., Schally A.V.;
 RT "Structure of the porcine LH- and FSH-releasing hormone. II.
 RT Confirmation of the proposed structure by conventional sequential
 RT analyses.";
 RL Biochem. Biophys. Res. Commun. 44:459-463 (1971).
 RN [3]
 RN SYNTHESIS OF GONADOLIBERIN.
 RX MEDLINE=72065376; PubMed=4942726;
 RA Matsuo H., Arimura A., Nair R.M.G., Schally A.V.;
 RT "Synthesis of the porcine LH- and FSH-releasing hormone by the solid-
 RT phase method.";
 RL Biochem. Biophys. Res. Commun. 45:822-827 (1971).
 RN [4]
 RN SYNTHESIS OF GONADOLIBERIN.
 RX MEDLINE=72117544; PubMed=4946275;
 RA Baba Y., Arimura A., Schally A.V.;
 RT "On the tryptophan residue in porcine LH and FSH-releasing hormone.";
 RL Biochem. Biophys. Res. Commun. 45:483-487 (1971).
 CC -1- FUNCTION: Stimulates the secretion of gonadotropins; it stimulates
 CC the secretion of both luteinizing and follicle-stimulating
 CC hormones.
 CC -1- SUBCELLULAR LOCATION: Secreted.
 CC -1- SIMILARITY: Belongs to the GnRH family.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; L32864; AAA31066.1; -;
 CC InterPro: IPR002012; GnRH.
 CC InterPro: IPR004079; Gonadoliberin1.
 CC Pfam: PF00446; GnRH; 1.
 CC PRINTS: PR01541; GONADOLIBERNI.
 CC PROSITE; PS00473; GnRH; 1.
 CC Cleavage on pair of basic residues; Hormone; Amidation; Hypothalamus;
 CC Placenta; Signal; Pyroglutamate carboxylic acid.
 FT SIGNAL 1 23
 FT CHAIN 24 91
 FT PEPTIDE 24 33
 FT PEPTIDE 34 91
 FT ACT_SITE 26 26
 FT MOD_RES 24 24
 FT MOD_RES 33 33
 FT SEQUENCE 91 AA; 10090 MW; 8340474F3DDAA99 CRC64;
 Query Match 96.3%; Score 52; DB 1; Length 91;
 Best Local Similarity 100.0%; Pred. No. 0.005;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 2 HWSYGLRP 9
 DB 25 HWSYGLRP 32

Query Match 96.3%; Score 52; DB 1; Length 67;
 Best Local Similarity 100.0%; Pred. No. 0.0036;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 HWSYGLRP 9
 DB 7 HWSYGLRP 14

RESULT 4
 GONI_XENLA STANDARD; PRT; 89 AA.

ID GONI_XENLA
 AC P45656;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Gonadoliberin I precursor (Gonadotropin-releasing hormone I) (GNRH-I)
 DE (LH-RH) (Luliberin I).
 OS Xenopus laevis (African clawed frog).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae;
 OC Xenopodinae; Xenopus.
 CX NCBI_TaxID=8355;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Forebrain;
 RX MEDLINE=94185563; PubMed=8137750;
 RA Hayes W.P., Wray S., Battley J.F.;
 RT "The frog gonadotropin-releasing hormone-I (GNRH-I) gene has a
 mammalian-like expression pattern and conserved domains in
 GNRH-associated peptide, but brain onset is delayed until
 metamorphosis.";
 RL Endocrinology 134:1835-1844(1994).
 CC -!- FUNCTION: Stimulates the secretion of gonadotropins.
 CC -!- SUBCELLULAR LOCATION: Secreted.
 CC -!- SIMILARITY: Belongs to the GNRH family.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).

EMBL; L28040; AAA49728.1;
 PIR; I51423; I51423.
 DR InterPro; IPR002012; GNRH.
 DR Pfam; PF00446; GNRH.1.
 DR PRINTS; PR01541; GONADOLIBERNI.
 DR PROSITE; PS00473; GNRH.1.
 DR Cleavage on pair of basic residues; Hormone; Amidation; Hypothalamus;
 KW Signal; Pyrrolidone carboxylic acid.
 FT SIGNAL 1 23
 FT CHAIN 24 89
 FT PEPTIDE 24 33
 FT PEPTIDE 37 89
 FT PEPTIDE 37 85
 FT PEPTIDE 24 24
 FT MOD_RES 24 24
 FT MOD_RES 33 33
 SQ SEQUENCE 89 AA; 10246 MW; 6F4F36FBAE0D4284 CRC64;

Query Match 96.3%; Score 52; DB 1; Length 89;
 Best Local Similarity 100.0%; Pred. No. 0.0049;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 HWSYGLRP 9
 DB 25 HWSYGLRP 32

RESULT 6
 GONI_RANCA STANDARD; PRT; 90 AA.

ID GONI_RANCA
 AC Q90Y63;
 DT 10-OCT-2003 (Rel. 42, Created)
 DT 10-OCT-2003 (Rel. 42, Last sequence update)

RESULT 5

GONI_MOUSE STANDARD; PRT; 90 AA.

ID GONI_MOUSE
 AC P13562;
 DT 01-JAN-1990 (Rel. 13, Created)
 DT 01-JAN-1990 (Rel. 13, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Progonadoliberin I precursor (Contains: Gonadoliberin I (LH-RH I)
 DE (Luteinizing hormone-releasing hormone I) (Gonadotropin-releasing
 DE hormone I) (GNRH I) (Luliberin I); Prolactin release-inhibiting factor
 DE I).
 GN GNRH1 OR GNRH.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 CX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=87069928; PubMed=3024317;
 RA Mason A.J., Hayflick J.S., Zoeller R.T., Young W.S. III,
 RA Phillips H.S., Nikolics K., Seeburg P.H.;
 RT "A deletion truncating the gonadotropin-releasing hormone gene is
 responsible for hypogonadism in the hpg mouse.";
 RL Science 234:1366-1371(1986).
 CC -!- FUNCTION: Stimulates the secretion of gonadotropins; it stimulates
 the secretion of both luteinizing and follicle-stimulating
 hormones.
 CC -!- SUBCELLULAR LOCATION: Secreted.
 CC -!- SIMILARITY: Belongs to the GNRH family.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).

EMBL; M14872; AAA37717.1;
 PIR; A47578; RHMSG.
 DR MGD; MGI:95789; Gnrh.
 DR InterPro; IPR002012; GNRH.
 DR InterPro; IPR004079; GonadoliberinI.
 DR Pfam; PF00446; GNRH.1.
 DR PRINTS; PR01541; GONADOLIBERNI.
 DR PROSITE; PS00473; GNRH.1.
 DR Cleavage on pair of basic residues; Hormone; Amidation; Hypothalamus;
 KW Placenta; Signal; Pyrrolidone carboxylic acid.
 FT SIGNAL 1 21
 FT CHAIN 22 90
 FT PEPTIDE 22 31
 FT PEPTIDE 35 90
 FT ACT_SITE 24 24
 FT ACT_SITE 24 24
 FT MOD_RES 22 22
 FT MOD_RES 31 31
 SQ SEQUENCE 90 AA; 10337 MW; 1C0766FA4826E4D9 CRC64;

Query Match 96.3%; Score 52; DB 1; Length 90;
 Best Local Similarity 100.0%; Pred. No. 0.0049;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 HWSYGLRP 9
 DB 23 HWSYGLRP 30

RESULT 6
 GONI_RANCA STANDARD; PRT; 90 AA.

ID GONI_RANCA
 AC Q90Y63;
 DT 10-OCT-2003 (Rel. 42, Created)
 DT 10-OCT-2003 (Rel. 42, Last sequence update)

Query Match 96.3%; Score 52; DB 1; Length 63;
Best Local Similarity 100.0%; Pred. No. 0.0034;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 HWSYGLRP 9
DB 2 HWSYGLRP 9

RESULT 3
GONI_MESAU STANDARD; PRT; 63 AA.

AC 009123;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Progonadoliberein I precursor [Contains: Gonadoliberein I (LH-RH I)
DE (Luteinizing hormone-releasing hormone I) (Gonadotropin-releasing
DE hormone I) (GnRH I) (Luliberin I); GnRH-associated peptide I]
DE (Fragment).
GN GNRH1 OR GNRH OR LHRH.
OS Macaca mulatta (Rhesus macaque).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
OC Cercopithecoidea; Macaca.
OX NCBI_TaxID=9544;
RN [1]
RP SEQUENCE FROM N.A.
RX TISSUE=Hypothalamus;
RA Ma Y.J., Costa M.E., Ojeda S.R.;
RT "Developmental expression of the genes encoding transforming growth
RT factor alpha and its receptor in the hypothalamus of female rhesus
RT macaques";
RL Neuroendocrinology 60:346-359(1994).
CC -!- FUNCTION: Stimulates the secretion of gonadotropins; it stimulates
CC the secretion of both luteinizing and follicle-stimulating
CC hormones.
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- SIMILARITY: Belongs to the GnRH family.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation
CC at the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC EMBL; U91938; AAB51302.1; -
DR InterPro; IPR002012; GnRH.
DR InterPro; IPR004079; GonadolibereinI.
DR Pfam; PF00446; GnRH; 1.
DR PRINTS; PR01541; GONADOLIBERNI.
DR PROSITE; PS00473; GnRH; 1.
KW Cleavage on pair of basic residues; Hormone; Amidation; Hypothalamus;
KW Placenta; Pyroglutamate carboxylic acid.
FT NON_TER 1 1
FT SIGNAL <1 5 BY SIMILARITY.
FT CHAIN 6 >67 PROGNADOLIBERIN I.
FT PEPTIDE 6 15 GONADOLIBERIN I.
FT ACT_SITE 19 >67 GnRH-ASSOCIATED PEPTIDE I.
FT MOD_RES 8 8 APPENDS TO BE ESSENTIAL FOR BIOLOGICAL
FT ACT_SITE 6 6 ACTIVITY (BY SIMILARITY).
FT MOD_RES 15 15 PYROGLUTAMATE CARBOXYLIC ACID (BY
FT MOD_RES 15 15 SIMILARITY).
FT NON_TER 67 67 AMIDATION (G-16 PROVIDE AMIDE GROUP) (BY
FT SEQUENCE 67 AA; 7573 MW; 505394DAA261A3F2 CRC64;

FT PEPTIDE 14 >61 GNRH-ASSOCIATED PEPTIDE I.
FT ACT_SITE 3 3 APPEARS TO BE ESSENTIAL FOR BIOLOGICAL
FT MOD_RES 1 1 ACTIVITY.
FT MOD_RES 10 10 PYROGLUTAMATE CARBOXYLIC ACID.
FT NON_TER 61 61 AMIDATION (G-11 PROVIDE AMIDE GROUP).
SQ SEQUENCE 61 AA; 6828 MW; 63962A1AB319B8F0 CRC64;

Query Match 96.3%; Score 52; DB 1; Length 61;
Best Local Similarity 100.0%; Pred. No. 0.0033;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 HWSYGLRP 9
DB 2 HWSYGLRP 9

RESULT 2
GONI_MESAU STANDARD; PRT; 63 AA.

AC 009123;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Progonadoliberein I precursor [Contains: Gonadoliberein I (LH-RH I)
DE (Luteinizing hormone-releasing hormone I) (Gonadotropin-releasing
DE hormone I) (GnRH I) (Luliberin I); GnRH-associated peptide I]
DE (Fragment).
GN GNRH1 OR GNRH OR LHRH.
OS Mesocricetus auratus (Golden hamster).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetidae;
OC Mesocricetus.
OX NCBI_TaxID=10036;
RN [1]
RP SEQUENCE FROM N.A.
RX Jansen H.T., Stevens P.J., Zeitler P., Lehman M.N.;
RA Submitted (MAR-1997) to the EMBL/GenBank/DBJ databases.
RL -!- FUNCTION: Stimulates the secretion of gonadotropins; it stimulates
RL the secretion of both luteinizing and follicle-stimulating
RL hormones.
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- SIMILARITY: Belongs to the GnRH family.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation
CC at the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC EMBL; U91938; AAB51302.1; -
DR InterPro; IPR002012; GnRH.
DR InterPro; IPR004079; GonadolibereinI.
DR Pfam; PF00446; GnRH; 1.
DR PRINTS; PR01541; GONADOLIBERNI.
DR PROSITE; PS00473; GnRH; 1.
KW Cleavage on pair of basic residues; Hormone; Amidation; Hypothalamus;
KW Placenta; Pyroglutamate carboxylic acid.
FT NON_TER 1 1
FT CHAIN 1 >63 PROGNADOLIBERIN I.
FT PEPTIDE 1 10 GONADOLIBERIN I.
FT ACT_SITE 14 >63 GnRH-ASSOCIATED PEPTIDE I (BY
FT MOD_RES 3 3 SIMILARITY).
FT MOD_RES 1 1 APPENDS TO BE ESSENTIAL FOR BIOLOGICAL
FT MOD_RES 10 10 ACTIVITY (BY SIMILARITY).
FT MOD_RES 10 10 PYROGLUTAMATE CARBOXYLIC ACID (BY
FT MOD_RES 63 63 SIMILARITY).
FT NON_TER 63 63 AMIDATION (G-11 PROVIDE AMIDE GROUP) (BY
FT SEQUENCE 63 AA; 7370 MW; FC9499567677180 CRC64;

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 10, 2004, 08:58:53 ; Search time 1.82879 Seconds
(without alignments)
284.724 Million cell updates/sec

Title: US-09-848-834A-1

Perfect score: 54

Sequence: 1 XHWSYGLRPX 10

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_42:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	52	96.3	61	GON1_SHEEP	Q28588 ovis aries
2	52	96.3	63	GON1_MESAU	O09163 mesocricetu
3	52	96.3	67	GON1_MACMU	P55247 macaca mula
4	52	96.3	89	GON1_XENLA	P45656 xenopus lae
5	52	96.3	90	GON1_MOUSE	P13562 mus musculus
6	52	96.3	90	GON1_RANCA	Q90763 rana catesb
7	52	96.3	91	GON1_PIG	P49921 sus scrofa
8	52	96.3	92	GON1_HUMAN	P01148 homo sapien
9	52	96.3	92	GON1_RAT	P07490 rattus norv
10	52	96.3	92	GON1_TUPGB	Q95335 tupaia glis
11	48	88.9	10	GON1_ALLMI	P37041 alligator m
12	48	88.9	92	GON1_CHICK	P37042 gallus gall
13	46	85.2	94	GON1_HAPBU	P51918 haplochromi
14	46	85.2	95	GON1_MORSA	O73812 morone saxa
15	46	85.2	95	GON1_PAGMA	P70074 pagrus majo
16	46	85.2	95	GON1_SPAAU	P51919 sparus majo
17	46	85.2	99	GON1_DICLA	Q91a10 dicentrarch
18	44	81.5	90	GON8_RANDY	Q91a02 rana dybows
19	43	79.6	92	GON1_CAVPO	O54713 cavia porce
20	42	77.8	80	GON1_CLAGA	P34339 claritas gar
21	42	77.8	91	GON1_ORYLA	Q96908 o progonado
22	41	75.9	10	GON1_CLUPA	P81749 clupea pall
23	41	75.9	110	YHEU_ACTAC	P96769 actinobacil
24	39	72.2	10	GON3_ONCKE	P20367 oncorhynch
25	39	72.2	74	GON3_ONCMY	P55246 oncorhynch
26	39	72.2	74	GON3_ONCTS	Q92097 oncorhynch
27	39	72.2	82	GON3_ONCMA	P30973 oncorhynch
28	39	72.2	82	GON3_SALSA	P35629 salmo salar
29	39	72.2	89	GON3_SALTU	P45653 salmo trutt
30	39	72.2	89	GON3_PORNO	P51922 porichthys
31	39	72.2	90	GON3_DICLA	Q91a09 dicentrarch
32	39	72.2	90	GON3_HAPBU	P45652 haplochromi
33	39	72.2	90	GON3_ORYLA	Q96d49 o progonado

34 39 72.2 90 1 GON3_PAGMA
35 39 72.2 90 1 GON3_SPAAU
36 39 72.2 94 1 GON3_CARAU
37 39 72.2 94 1 GON3_RUTFU
38 36 66.7 437 1 CAH9_MOUSE
39 35 64.8 82 1 YF76_ARCFU
40 35 64.8 315 1 ALD2_MOUSE
41 35 64.8 364 1 ARGCC_BIFLO
42 35 64.8 371 1 YLII_ECOLI
43 35 64.8 473 1 C3AR_RAT
44 35 64.8 477 1 C3AR_MOUSE
45 35 64.8 584 1 GUND_CLOCE

ALIGNMENTS

RESULT 1
GON1_SHEEP
ID GON1_SHEEP STANDARD; PRT; 61 AA.
AC Q28588
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Progonadoliberin I precursor [Contains: Gonadoliberin I (LH-RH I)
(Luteinizing hormone-releasing hormone I) (Gonadotropin-releasing
hormone I) (GnRH I) (Luliberin I); GnRH-associated peptide I]
DE (Fragment)
DE GN RH I OR GN RH OR LHRH.
OS Ovis aries (Sheep).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Caprinae; Ovis.
OX NCBI_TaxID=9940;
RN [1]
RP SEQUENCE OF 12-61 FROM N.A.
RC STRAIN=Western range; TISSUE=Hypothalamus;
RA Rodriguez R.E.; Wise M.E.;
RL Submitted (OCT-1993) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE OF 1-10.
RX MEDLINE=72094314; PubMed=4550508;
RA Burkus R., Butcher M., Amoss M., Ling N., Monahan M., Rivier J.,
Fellows R., Blackwell R., Vale W., Guillemin R.;
RT "Primary structure of the ovine hypothalamic luteinizing hormone-
releasing factor (LRF) (LH-hypothalamus-LRF-gas chromatography-mass
spectrometry-decapeptide-Edman degradation).";
RL Proc. Natl. Acad. Sci. U.S.A. 69:278-282(1972).
CC -!- FUNCTION: Stimulates the secretion of gonadotropins; it stimulates
the secretion of both luteinizing and follicle-stimulating
hormones.
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- SIMILARITY: Belongs to the GnRH family.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@sib-sib.ch).
CC EMBL; U02517; AAA03433.1;
CC InterPro; IPR002012; GnRH
CC InterPro; IPR004079; Gonadoliberin I.
CC Pfam; PF00446; GnRH; 1.
CC PRINTS; PR01541; GONADOLIBERNI.
CC PROSITE; PS00473; GnRH; 1.
CC Cleavage on pair of basic residues; Hormone; Amidation; Hypothalamus;
CC Placenta; Pyrrolidone carboxylic acid.
CC NON TER 1 1
CHAIN 1 >61 PROGONADOLIBERIN I.
PEPTIDE 1 10 GONADOLIBERIN I.

```
CC -----
DR EMBL; U00062; AAB68915.1; -.
DR PIR; S46746; S46746.
DR GenOnline; 139356; -.
DR SGD; S0001081; MSC7.
DR GO; GO:0007131; P:meiotic recombination; IMP.
DR InterPro; IPR02086; Aldehyde_dehydr.
DR Pfam; PF00171; aldedh; 1.
DR PROSITE; PS00070; ALDEHYDE DEHYDR_CYS; 1.
DR PROSITE; PS00687; ALDEHYDE DEHYDR_GLU; 1.
KW Hypothetical protein; Oxidoreductase.
FT ACT_SITE 354 354 BY SIMILARITY.
FT ACT_SITE 389 389 BY SIMILARITY.
SQ SEQUENCE 644 AA; 71320 MW; 54DADDAEB2A16D4D CRC64;

Query Match 50.6%; Score 40; DB 1; Length 644;
Best Local Similarity 60.0%; Pred. No. 18;
Matches 9; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 2 QYIKANSKFIGITEL 16
Db 38 QIQDNQKLIGITTL 52

Search completed: March 10, 2004, 09:13:47
Job time : 4.92607 secs
```

```

RT "Biochemical and genetic characterization of PepF, an oligopeptidase
RT from Lactococcus lactis."
RT J. Biol. Chem. 269:32070-32076 (1994).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=NCDO 763;
RX MEDLINE=97352670; PubMed=9209029;
RA Nardi M., Renault P., Monnet V.;
RT "Duplication of the pepF gene and shuffling of DNA fragments on the
RT lactose plasmid of Lactococcus lactis."
RT J. Bacteriol. 179:4164-4171 (1997).
RL Lactococcus lactis
CC -1- FUNCTION: Hydrolyzes peptides containing between 7 and 17 amino
CC acids with a rather wide specificity.
CC -1- COFACTOR: Binds 1 zinc ion (By similarity).
CC -1- SIMILARITY: Belongs to peptidase family M3.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; Z32522; CAA83534.1; -
CC EMBL; X99798; CAA68133.1; -
CC PIR; A55485; A55485.
CC MEROPS; M03.007; -
CC InterPro; IPR006025; Pept_M_Zn_BS.
CC InterPro; IPR001567; Peptidase_M3.
CC InterPro; IPR004438; Peptidase_M3B.
CC Pfam; PF01432; Peptidase_M3; 1.
CC TIGRfam; TIGR00181; pepF; 1.
CC PROSITE; PS00142; ZINC_PROTEASE; FALSE NEG.
CC Hydrolase; Metalloprotease; Zinc; Plasmid.
FT METAL 387 388 ZINC (CATALYTIC) (BY SIMILARITY).
FT ACT SITE 388 388 BY SIMILARITY.
FT METAL 391 391 ZINC (CATALYTIC) (BY SIMILARITY).
FT METAL 394 394 ZINC (CATALYTIC) (BY SIMILARITY).
FT CONFLICT 518 518 F -> S (IN REF. 2).
SQ SEQUENCE 601 AA; 69674 MW; C9B5C519FFA2F787 CRC64;

Query Match 50.6%; Score 40; DB 1; Length 601;
Best Local Similarity 46.7%; Pred. NO. 17;
Matches 7; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 2 QVYKANSKFIGITEL 16
DB 284 RYIELRKKILGITDL 298

RESULT 14
PEPF LACLA STANDARD; PRT; 601 AA.
ID Q9CSV7;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Oligoendopeptidase F homolog (EC 3.4.24.-).
GN PEPF OR LL1727.
OS Lactococcus lactis (subsp. lactis) (Streptococcus lactis).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae; Lactococcus.
OX NCBI_TaxID=1360;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=IL1403;
RX MEDLINE=21235186; PubMed=11337471;
RA Bolotin A., Wincker P., Mauger S., Jaillon O., Malarne K.,
RA Weissenbach J., Ehrlich S.D., Sorokin A.;
RT "The complete genome sequence of the lactic acid bacterium Lactococcus
RT lactis ssp. lactis IL1403."
RL Genome Res. 11:731-753 (2001).
CC -1- FUNCTION: Hydrolyzes peptides containing between 7 and 17 amino

```

```

CC acids with a rather wide specificity (By similarity).
CC -1- COFACTOR: Binds 1 zinc ion (By similarity).
CC -1- SIMILARITY: Belongs to peptidase family M3.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AE006403; AAK05825.1; -
CC PIR; G86840; G86840.
CC InterPro; IPR006025; Pept_M_Zn_BS.
CC InterPro; IPR001567; Peptidase_M3.
CC InterPro; IPR004438; Peptidase_M3B.
CC Pfam; PF01432; Peptidase_M3; 1.
CC TIGRfam; TIGR00181; pepF; 1.
CC PROSITE; PS00142; ZINC_PROTEASE; 1.
CC Hydrolase; Metalloprotease; Zinc; Complete proteome.
FT METAL 387 388 ZINC (CATALYTIC) (BY SIMILARITY).
FT ACT SITE 388 388 BY SIMILARITY.
FT METAL 391 391 ZINC (CATALYTIC) (BY SIMILARITY).
FT METAL 394 394 ZINC (CATALYTIC) (BY SIMILARITY).
SQ SEQUENCE 601 AA; 69883 MW; 045E6873F10FE19A CRC64;

Query Match 50.6%; Score 40; DB 1; Length 601;
Best Local Similarity 46.7%; Pred. NO. 17;
Matches 7; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 2 QVYKANSKFIGITEL 16
DB 284 RYIELRKKILGITDL 298

RESULT 15
YHJ9 YEAST STANDARD; PRT; 644 AA.
ID YHJ9 YEAST
AC P38694;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Hypothetical aldehyde-dehydrogenase like protein in FILL-VMA10
DE intergenic region (EC 1.2.1.-).
GN YHJ039C.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S288c / AB972;
RX MEDLINE=94378003; PubMed=8091229;
RA Johnston M., Andrews S., Brinkman R., Cooper J., Ding H., Dover J.,
Du Z., Favello A., Fulton L., Gattung S., Geisel C., Kirsten J.,
Kucaba T., Hillier L.W., Jier M., Johnston L., Langston Y.,
Latreille P., Louis E.J., Macri C., Mardis E., Menezes S., Mouser L.,
Nhan M., Rifkin L., Riles L., St Peter H., Trevasakis E., Vaughan K.,
Vignati D., Wilcox L., Wohlman P., Waterston R., Wilson R.,
Vaughan M.;
RT "Complete nucleotide sequence of Saccharomyces cerevisiae chromosome
RT VIII."
RL Science 265:2077-2082 (1994).
CC -1- SIMILARITY: Belongs to the aldehyde dehydrogenase family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).

```

AC P43013;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Probable acyl carrier protein phosphodiesterase (EC 3.1.4.14) (ACP
 DE phosphodiesterase).
 GN ACPD OR H11366.
 OS Haemophilus influenzae.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
 OC Pasteurellaceae; Haemophilus.
 OX NCBI_TaxID=727;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Rd / KW20 / ATCC 51907;
 EX MEDLINE=96186898; PubMed=8635745;
 RA Chandler M.S., Smith R.A.;
 RT Characterization of the Haemophilus influenzae topA locus: DNA
 RT topoisomerase I is required for genetic competence.";
 RL Gene 169:25-31(1996).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Rd / KW20 / ATCC 51907;
 EX MEDLINE=95350630; PubMed=7542800;
 RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
 RA Kierlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
 RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
 RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
 RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom B., Cotton M.D.,
 RA Uterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
 RA Fine L.D., Fitchman J.L., Fuhmann J.L., Geoghagen N.S.M.,
 RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
 RA Venter J.C.;
 RT "Whole-genome random sequencing and assembly of Haemophilus influenzae
 RT Rd.";
 RL Science 269:496-512(1995).
 CC -!- FUNCTION: Converts holo-ACP to apo-ACP by hydrolytic cleavage of
 CC the phosphopantetheine prosthetic group from ACP (By similarity).
 CC -!- CATALYTIC ACTIVITY: Holo-[acyl-carrier protein] + H(2)O = 4'-
 CC phosphopantetheine + apo-[acyl-carrier protein].
 CC -!- SIMILARITY: Belongs to the acpD family.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; U20964; AAC43728.1; -;
 DR EMBL; U32816; AAC23013.1; -;
 DR PIR; G64026; G64026.
 DR TIGR; H11366; -;
 DR HAMAP; MF 01216; -; 1.
 DR InterPro; IPR003680; NADPHd 2.
 DR Pfam; PF02525; Flavodoxin 2; 1.
 KW Hydrolase; Complete proteome.
 SQ SEQUENCE 194 AA; 21208 MW; A8AEC9D00829522C CRC64;
 Query Match 50.6%; Score 40; DB 1; Length 194;
 Best Local Similarity 53.3%; Pred. No. 5.4;
 Matches 8; Conservative 4; Mismatches 3; Indels 0; Gaps 0;
 QY 2 QVIRKANSKFIGITEL 16
 DB 147 QYMKSLGFIGITDV 161
 RESULT 12
 KAD_MYCCA
 ID KAD_MYCCA STANDARD; PRT; 213 AA.
 AC P10251;
 DT 01-MAR-1989 (Rel. 10, Created)

DT 01-MAR-1989 (Rel. 10, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Adenylate kinase (EC 2.7.4.3) (ATP-AMP transphosphorylase).
 GN ADK.
 OS Mycoplasma capricolum.
 OC Bacteria; Firmicutes; Mollicutes; Mycoplasmatataceae; Mycoplasma.
 OX NCBI_TaxID=2095;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 27343 / Kid;
 RX MEDLINE=88142549; PubMed=3481422;
 RA Ohkubo S., Muto A., Kawauchi Y., Yamao F., Osawa S.;
 RT "The ribosomal protein gene cluster of Mycoplasma capricolum.";
 RL Mol. Gen. Genet. 210:314-322(1987).
 CC -!- FUNCTION: This small ubiquitous enzyme is essential for
 CC maintenance and cell growth.
 CC -!- CATALYTIC ACTIVITY: ATP + AMP = ADP + ADP.
 CC -!- SUBUNIT: Monomer (By similarity).
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
 CC -!- SIMILARITY: Belongs to the adenylate kinase family.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; X06414; CAA29724.1; -;
 DR EMBL; X06414; CAA29724.1; -;
 DR PIR; S02851; KIYMC.
 DR HSP; P27142; LZIN.
 DR HAMAP; MF 00235; -; 1.
 DR InterPro; IPR006259; Adenyl_kin sub.
 DR InterPro; IPR000850; Adenylate_Kin.
 DR InterPro; IPR007882; ADK_lid.
 DR Pfam; PF0406; ADK; 1.
 DR Pfam; PF05191; ADK lid; 1.
 DR PRINTS; PR00094; ADENYLKINASE.
 DR PRODOM; PD000657; Adenylate kin; 1.
 DR TIGRfam; TIGR01351; adk; 1.
 DR PROSITE; PS00113; ADENYLATE KINASE; 1.
 KW Transferase; Kinase; ATP-binding.
 FT NP_BIND 7 15 ATP (BY SIMILARITY).
 SQ SEQUENCE 213 AA; 24616 MW; FBCFA426B6F92E16 CRC64;
 Query Match 50.6%; Score 40; DB 1; Length 213;
 Best Local Similarity 61.5%; Pred. No. 6;
 Matches 8; Conservative 1; Mismatches 4; Indels 0; Gaps 0;
 QY 1 MOYIKANSKFIGI 13
 DB 181 IDYFKTNSKFIEI 193
 RESULT 13
 PEPI_LACILC
 ID PEPI_LACILC STANDARD; PRT; 601 AA.
 AC P54124; P94880;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Oligonucleotide phosphatase P, plasmid (EC 3.4.24.-).
 GN PEPI OR PEPP
 OS Lactococcus lactis (subsp. cremoris) (Streptococcus cremoris).
 OG Plasmid pUP763.
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae; Lactococcus.
 OX NCBI_TaxID=1359;
 RN [1]
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
 RC STRAIN=NCDO 763;
 RX MEDLINE=95096044; PubMed=7798200;
 RN Monnet V., Nardi M., Chopin A., Chopin M.-C., Gripon J.-C.;

Wed Mar 10 10:34:31 2004

us-09-848-834a-2.open.rsp

use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announcement/> or send an email to license@isb-sib.ch).

EMBL; AF175466; AAD50307.1; ALT_INIT.
DR HAMAP; MF_00219; -; 1.
DR InterPro; IPR006680; Amidohydro_1.
DR InterPro; IPR002195; Pept_M38_nph.
DR Pfam; PF01979; Amidohydro_1; 1.
DR PROSITE; PS00482; DIHYDROOROTASE 1; PARTIAL.
DR PROSITE; PS00483; DIHYDROOROTASE 2; 1.
KW Pyrimidine biosynthesis; Hydroxylase; Metal-binding; Zinc.
FT NON_TER 1
FT METAL 34 34 ZINC 2 (BY SIMILARITY).
FT METAL 107 107 ZINC 1 (BY SIMILARITY).
FT SEQUENCE 204 AA; 22899 MW; 9DB0B6C9B834B310 CRC64;

Query Match 51.9%; Score 41; DB 1; Length 204;
Best Local Similarity 46.2%; Pred. No. 3; 7;
Matches 6; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

QY 2 QYKANSKFIGIT 14
[:::]:[:]
Db 42 QYVOAGNRFGLGAT 54

RESULT 9
LE12_THETN STANDARD; PRT; 384 AA.
ID LE12_THETN
AC Q8RCF9;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE 2-isopropylmalate synthase 2 (EC 2.3.3.13) (Alpha-isopropylmalate synthase 2) (Alpha-IPM synthetase 2).
GN LEUA2 OR TTE0472.
OS Thermoanaerobacter tengcongensis.
OC Bacteria; Firmicutes; Clostridia; Thermoanaerobacteriales;
OC Thermoanaerobacteriaceae; Thermoanaerobacter.
OX NCBI_TaxID=119072;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MB4 / JCM 11007;
RX MEDLINE=21992816; PubMed=11997336;
RA Bao Q., Tian Y., Li W., Xu Z., Xuan Z., Hu S., Dong W., Yang J., Tan H., Chen Y., Xue Y., Xu Y., Lai X., Huang L., Dong X., Ma Y., Ling L., Tan H., Chen R., Wang J., Yu J., Yang H.;
RT "A complete sequence of T. tengcongensis genome.";
RL Genome Res. 12:689-700(2002).
CC -!- FUNCTION: Catalyzes the condensation of the acetyl group of acetyl-CoA with 3-methyl-2-oxobutanoate (2-oxoisovalerate) to form 3-carboxy-3-hydroxy-4-methylpentanoate (2-isopropylmalate).
CC -!- CATALYTIC ACTIVITY: Acetyl-CoA + 3-methyl-2-oxobutanoate + H(2)O = 2-hydroxy-2-isopropylsuccinate + CoA.
CC -!- PATHWAY: Leucine biosynthesis; first step.
CC -!- SUBUNIT: Homotrimer (By similarity).
CC -!- SIMILARITY: Belongs to the alpha-IPM synthetase / homocitrate synthase family. Leua 1 subfamily.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announcement/> or send an email to license@isb-sib.ch).

EMBL; AE013018; AAM23753.1; -.
DR HAMAP; MF_01025; atypical; 1.
DR InterPro; IPR002034; AIPW/Hcit_synth.
DR InterPro; IPR000891; HMGL-like.
DR Pfam; PF00682; HMGL-like; 1.

DR PROSITE; PS00815; AIPM_HOMOCIT_SYNTH_1; 1.
DR PROSITE; PS00816; AIPM_HOMOCIT_SYNTH_2; 1.
KW Leucine biosynthesis; Transferase; Complete proteome.
SQ SEQUENCE 384 AA; 42404 MW; 095310F2C0E4A4DD CRC64;

Query Match 51.9%; Score 41; DB 1; Length 384;
Best Local Similarity 54.5%; Pred. No. 7; 1;
Matches 6; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 4 IKANSKFIGIT 14
[:::]:[:]
Db 214 VKAGAEVGVGT 224

RESULT 10

VI20_HSVSA STANDARD; PRT; 899 AA.
ID VI20_HSVSA
AC Q01055;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE Capsid assembly protein 63.
GN 63 OR EEPF1.
OS Herpesvirus saimiri (strain 11).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Gammaherpesvirinae; Rhadinovirus.
OX NCBI_TaxID=10383;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92333688; PubMed=1321287;
RA Albrecht J.-C., Nicholas J., Biller D., Cameron K.R., Biesinger B., Newman C., Wittmann S., Craxton M.A., Coleman H., Fleckenstein B., Honess R.W.;
RT "Primary structure of the herpesvirus saimiri genome.";
RL J. Virol. 66:5047-5058(1992).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=92230228; PubMed=1314457;
RA Nicholas J., Cameron K.R., Coleman H., Newman C., Honess R.W.;
RT "Analysis of nucleotide sequence of the rightmost 43 kbp of herpesvirus saimiri (HVS) L-DNA: General conservation of genetic organization between HVS and Epstein-Barr virus.";
RL Virology 188:296-310(1992).
CC -!- SIMILARITY: BELONGS TO FAMILY THAT GROUPS TOGETHER HSV-1 UL37, EBV-1 23, EBV BOLF1, VZV 21, HVS-1 63, AND HCMV UL47.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announcement/> or send an email to license@isb-sib.ch).

EMBL; X64346; CAA45686.1; -.
DR EMBL; M86409; AAA46139.1; -.
DR InterPro; IPR008643; Herpes_ORF63.
DR Pfam; PF05765; Herpes_ORF63; 1.
KW Capsid assembly.
SQ SEQUENCE 899 AA; 103350 MW; F1429B3770A2885E CRC64;

Query Match 51.9%; Score 41; DB 1; Length 899;
Best Local Similarity 50.0%; Pred. No. 17;
Matches 7; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

QY 2 QYKANSKFIGIT 15
[:::]:[:]
Db 124 QYTSNATFTGLSE 137

RESULT 11
ACPD_HAEIN STANDARD; PRT; 194 AA.
ID ACPD_HAEIN

RA Weber J.L., Sim B.K.L., Lyon J.A., Wolff R.;
 RT "Merozoite surface protein sequence from the Camp strain of the human
 RL malaria parasite Plasmodium falciparum";
 RL Nucleic Acids Res. 16:1206-1206(1988).
 CC -1- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor
 CC (Potential).
 CC -1- PFM: Merozoite surface antigen contain the sequence of 83 kDa, 42
 CC kDa and 19 kDa antigens which are the major surface antigens of
 CC merozoites. The maturation take place during schizont.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; X03831; CA27446.1; -;
 DR PIR; A23386; SAZQGM.
 DR InterPro; IPR006209; EGF_like.
 DR Pfam; PF00008; EGF; 1.
 KW Malaria; Merozoite; Polyprotein; Repeat; Signal; Glycoprotein;
 KW Transmembrane; GPI-anchor.
 FT SIGNAL 1 19 POTENTIAL.
 FT CHAIN 20 1726 MEROZOITE SURFACE PROTEIN 1.
 FT CARBOHYD 133 133 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 272 272 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 501 501 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 567 567 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 567 567 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 638 638 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 827 827 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 924 924 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 944 944 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 990 990 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 1016 1016 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 1114 1114 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 1221 1221 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 1613 1613 N-LINKED (GLCNAC. .) (POTENTIAL).
 SQ SEQUENCE 1726 AA; 196197 MW; D8AD45FA352BCF3 CRC64;
 Query Match 53.8%; Score 42.5; DB 1; Length 1726;
 Best Local Similarity 60.0%; Pred. No. 17;
 Matches 9; Conservative 4; Mismatches 1; Indels 1; Gaps 1;
 QY 2 QYIKANSKFI-GITE 15
 Db 1026 QVKSNSKVIITGLTE 1040
 RESULT 7
 MSF1_PLAPP STANDARD; PRT; 1726 AA.
 AC F50495;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 01-OCT-1996 (Rel. 34, Last annotation update)
 DE Merozoite surface protein 1 precursor (Merozoite surface antigens)
 DE (PFMSA) (GP195).
 GN MSP-1.
 OS Plasmodium falciparum (isolate Palo Alto / Uganda).
 CC Eukaryota; Alveolata; Apicomplexa; Haemosporidia; Plasmodium.
 OX NCBI_TaxID=57270;
 RN [1] _
 RP SEQUENCE FROM N.A.
 RX MEDLINE=89005525; PubMed=3049134;
 RA Chang S.P., Kramer K.J., Yamaga K.M., Kato A., Case S.E.,
 RA Siddiqui W.A.;
 RT "Plasmodium falciparum: gene structure and hydropathy profile of the
 RT major merozoite surface antigen (gp195) of the Uganda-Palo Alto
 RT isolate.";
 RL Exp. Parasitol. 67:1-11(1988).
 CC -1- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor

CC (Potential).
 CC -1- PFM: Merozoite surface antigen contain the sequence of 83 kDa, 42
 CC kDa and 19 kDa antigens which are the major surface antigens of
 CC merozoites. The maturation take place during schizont.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; M37213; AAA29611.1; -;
 DR InterPro; IPR006209; EGF_like.
 DR Pfam; PF00008; EGF; 1.
 KW Malaria; Merozoite; Polyprotein; Repeat; Signal; Glycoprotein;
 KW Transmembrane; GPI-anchor.
 FT SIGNAL 1 19 POTENTIAL.
 FT CHAIN 20 1726 MEROZOITE SURFACE PROTEIN 1.
 FT CARBOHYD 133 133 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 272 272 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 501 501 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 567 567 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 567 567 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 638 638 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 827 827 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 924 924 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 944 944 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 990 990 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 1016 1016 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 1114 1114 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 1221 1221 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 1613 1613 N-LINKED (GLCNAC. .) (POTENTIAL).
 SQ SEQUENCE 1726 AA; 196174 MW; 5B59CEBFA2F9A026 CRC64;
 Query Match 53.8%; Score 42.5; DB 1; Length 1726;
 Best Local Similarity 60.0%; Pred. No. 17;
 Matches 9; Conservative 4; Mismatches 1; Indels 1; Gaps 1;
 QY 2 QYIKANSKFI-GITE 15
 Db 1026 QVKSNSKVIITGLTE 1040
 RESULT 8
 PYRC_SERMA STANDARD; PRT; 204 AA.
 ID Q9S3S1;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Dihydroorotase (EC 3.5.2.3) (DHOase) (Fragment).
 GN PYRC.
 OS Serratia marcescens.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Serratia.
 OX NCBI_TaxID=615;
 RN [1] _
 RP SEQUENCE FROM N.A.
 RC STRAIN=SM6;
 RA Berkmen M., Benedik M.J.;
 RT "Dini inhibits transcription of Serratia marcescens nuclease.";
 RL Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.
 CC -1- CATALYTIC ACTIVITY: (S)-dihydroorotate + H(2)O = N-carbamoyl-L-
 CC aspartate.
 CC -1- COFACTOR: Binds 2 zinc ions per subunit (By similarity).
 CC -1- PATHWAY: Pyrimidine biosynthesis; third step.
 CC -1- SUBUNIT: Homodimer (By similarity).
 CC -1- SIMILARITY: Belongs to the DHOase family. Subfamily 1.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its


```

CC  -!- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor
CC  (Potential).
CC  -!- PTM: Merozoite surface antigen contain the sequence of 83 kDa, 42
CC  kDa and 19 kDa antigens which are the major surface antigens of
CC  merozoites. The maturation take place during schizont.
CC  -----
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC  -----
CC  EMBL; M19143; AAA29653.1; -
CC  F1R; A54498; A54498.
CC  InterPro; IPR006209; EGF_like.
CC  Pfam; PF00008; EGF; 1.
CC  Malaria; Merozoite; GPI-anchor.
CC  Transmembrane; GPI-anchor.
CC  SIGNAL 1 19 POTENTIAL.
CC  CHAIN 20 1701 MEROZOITE SURFACE PROTEIN 1.
CC  CARBOHYD 110 110 N-LINKED (GLCNAC. .) (POTENTIAL).
CC  CARBOHYD 239 239 N-LINKED (GLCNAC. .) (POTENTIAL).
CC  CARBOHYD 470 470 N-LINKED (GLCNAC. .) (POTENTIAL).
CC  CARBOHYD 536 536 N-LINKED (GLCNAC. .) (POTENTIAL).
CC  CARBOHYD 607 607 N-LINKED (GLCNAC. .) (POTENTIAL).
CC  CARBOHYD 802 802 N-LINKED (GLCNAC. .) (POTENTIAL).
CC  CARBOHYD 899 899 N-LINKED (GLCNAC. .) (POTENTIAL).
CC  CARBOHYD 919 919 N-LINKED (GLCNAC. .) (POTENTIAL).
CC  CARBOHYD 965 965 N-LINKED (GLCNAC. .) (POTENTIAL).
CC  CARBOHYD 991 991 N-LINKED (GLCNAC. .) (POTENTIAL).
CC  CARBOHYD 1089 1089 N-LINKED (GLCNAC. .) (POTENTIAL).
CC  CARBOHYD 1196 1196 N-LINKED (GLCNAC. .) (POTENTIAL).
CC  CARBOHYD 1588 1588 N-LINKED (GLCNAC. .) (POTENTIAL).
CC  SEQUENCE 1701 AA; 193719 MW; 3920875E73D38552 CRC64;
CC
CC  Query Match 53.8%; Score 42.5; DB 1; Length 1701;
CC  Best Local Similarity 60.0%; Pred. No. 17;
CC  Matches 9; Conservative 4; Mismatches 1; Indels 1; Gaps 1;
CC
CC  QY 2 QYIKANSKEI-GITE 15
CC  |:::|::|::|::|
CC  Db 1001 QFVKSNSKVITGLTE 1015
CC
CC  RESULT 5
CC  MSPI_PLAFM STANDARD; PRT; 1701 AA.
CC  AC F08569;
CC  DT 01-AUG-1988 (Rel. 08, Created)
CC  DT 30-MAY-2000 (Rel. 39, Last sequence update)
CC  DT 30-MAY-2000 (Rel. 39, Last annotation update)
CC  DE Merozoite surface protein 1 precursor (Merozoite surface antigens)
CC  DE (PMWSA) (P190).
CC  GN MSP-1.
CC  OS Plasmodium falciparum (isolate mad20 / Papua New Guinea).
CC  OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
CC  OX NCBI_TaxID=70153;
CC  RN [1]
CC  RP SEQUENCE FROM N.A.
CC  RX MEDLINE=88011243; PubMed=3079521;
CC  RA Tanabe K., Mackay M., Goman M., Scaife J.G.;
CC  RT "Allelic dimorphism in a surface antigen gene of the malaria parasite
CC  Plasmodium falciparum."
CC  RL J. Mol. Biol. 195:273-287(1987).
CC  [2]
CC  RP REVISIONS TO 1403; 1569 AND 1629.
CC  RA Tanabe K.;
CC  RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.
CC  [3]
CC  RN SEQUENCE OF 1-115 FROM N.A.
CC  RP MEDLINE=86136024; PubMed=3004972;

```

```

RA Mackay M., Goman M., Bone N., Hyde J.E., Scaife J., Certa U.,
RA Stuenkel H., Bujard H.;
RA Polymorphism of the precursor for the major surface antigens of
RA Plasmodium falciparum merozoites: studies at the genetic level.;
RA EMBO J. 4:3823-3829(1985).
CC -!- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor
CC (Potential).
CC -!- PTM: Merozoite surface antigen contain the sequence of 83 kDa, 42
CC kDa and 19 kDa antigens which are the major surface antigens of
CC merozoites. The maturation take place during schizont.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X05624; CAA29112.1; -
CC InterPro; IPR006209; EGF_like.
CC Pfam; PF00008; EGF; 1.
CC Malaria; Merozoite; Polyprotein; Repeat; Signal; Glycoprotein;
CC Transmembrane; GPI-anchor.
CC SIGNAL 1 19 POTENTIAL.
CC CHAIN 20 1701 MEROZOITE SURFACE PROTEIN 1.
CC CARBOHYD 110 110 N-LINKED (GLCNAC. .) (POTENTIAL).
CC CARBOHYD 239 239 N-LINKED (GLCNAC. .) (POTENTIAL).
CC CARBOHYD 470 470 N-LINKED (GLCNAC. .) (POTENTIAL).
CC CARBOHYD 536 536 N-LINKED (GLCNAC. .) (POTENTIAL).
CC CARBOHYD 607 607 N-LINKED (GLCNAC. .) (POTENTIAL).
CC CARBOHYD 802 802 N-LINKED (GLCNAC. .) (POTENTIAL).
CC CARBOHYD 899 899 N-LINKED (GLCNAC. .) (POTENTIAL).
CC CARBOHYD 919 919 N-LINKED (GLCNAC. .) (POTENTIAL).
CC CARBOHYD 965 965 N-LINKED (GLCNAC. .) (POTENTIAL).
CC CARBOHYD 991 991 N-LINKED (GLCNAC. .) (POTENTIAL).
CC CARBOHYD 1089 1089 N-LINKED (GLCNAC. .) (POTENTIAL).
CC CARBOHYD 1196 1196 N-LINKED (GLCNAC. .) (POTENTIAL).
CC CARBOHYD 1588 1588 N-LINKED (GLCNAC. .) (POTENTIAL).
CC SEQUENCE 1701 AA; 193768 MW; 3FC2EC59AF96BA98 CRC64;
CC
CC  Query Match 53.8%; Score 42.5; DB 1; Length 1701;
CC  Best Local Similarity 60.0%; Pred. No. 17;
CC  Matches 9; Conservative 4; Mismatches 1; Indels 1; Gaps 1;
CC
CC  QY 2 QYIKANSKEI-GITE 15
CC  |:::|::|::|::|
CC  Db 1001 QFVKSNSKVITGLTE 1015
CC
CC  RESULT 6
CC  MSPI_PLAFM STANDARD; PRT; 1726 AA.
CC  AC P04934;
CC  DT 13-AUG-1987 (Rel. 05, Created)
CC  DT 01-MAR-1989 (Rel. 10, Last sequence update)
CC  DT 15-MAR-2004 (Rel. 43, Last annotation update)
CC  DE Merozoite surface protein 1 precursor (Merozoite surface antigens)
CC  DE (PMWSA) (P195).
CC  GN MSP-1.
CC  OS Plasmodium falciparum (isolate Camp / Malaysia).
CC  OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
CC  OX NCBI_TaxID=5835;
CC  RN [1]
CC  RP SEQUENCE OF 1-1103 FROM N.A.
CC  RX MEDLINE=86205236; PubMed=3517809;
CC  RA Weber J.L., Leininger W.M., Lyon J.A.;
CC  RT "Variation in the gene encoding a major merozoite surface antigen of
CC  the human malaria parasite Plasmodium falciparum."
CC  RL Nucleic Acids Res. 14:3311-3323(1986).
CC  [2]
CC  RN SEQUENCE OF 1104-1726 FROM N.A.
CC  RP MEDLINE=88143999; PubMed=3278296;

```

Query Match 100.0%; Score 79; DB 1; Length 1314;
Best Local Similarity 100.0%; Pred. No. 2.7e-06;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MOYIKANSKFIGITEL 16
Db 828 MOYIKANSKFIGITEL 843

RESULT 2

VG84 BPML5 STANDARD; PRT; 66 AA.
AC Q05301; 1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 01-FEB-1994 (Rel. 28, Last annotation update)
DE Gene 84 protein (GP84).
GN 84.
OS Mycobacteriophage L5.
OC Viruses; dsDNA viruses, no RNA stage; Caudovirales; Siphoviridae;
OC L5-like viruses.
OX NCBI_TaxID=311757;
RN [1]

SEQUENCE FROM N.A.
RX MEDLINE=93211282; PubMed=8459766;
RA Hatfull G.F., Sarkis G.J.;
RT "DNA sequence, structure and gene expression of mycobacteriophage L5;
RL Mol. Microbiol. 7:395-405(1993).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M35727; AAA29715.1; -;
DR EMBL; Y00087; CAA68280.1; -;
DR EMBL; Z35326; CAA84555.1; -;
DR InterPro: IPR006209; EGF_like.
DR Pfam: PF00008; EGF; 1.
KW Malaria; Merozoite; GPI-anchored; Polyprotein; Repeat; Signal; Glycoprotein;
KW Transmembrane; GPI-anchor.
FT SIGNAL 1 19
FT CHAIN 20 1682
FT TRANSMEM 1666 1682
FT CARBOHYD 233 233
FT CARBOHYD 462 462
FT CARBOHYD 528 528
FT CARBOHYD 599 599
FT CARBOHYD 785 785
FT CARBOHYD 881 881
FT CARBOHYD 901 901
FT CARBOHYD 947 947
FT CARBOHYD 1071 1071
FT CARBOHYD 1178 1178
FT CARBOHYD 1569 1569
FT CARBOHYD 1682 1682
SQ SEQUENCE 1682 AA; 192462 MW; C82A1E159948CAD6 CRC64;

Query Match 53.8%; Score 42.5; DB 1; Length 1682;
Best Local Similarity 60.0%; Pred. No. 17;
Matches 9; Conservative 4; Mismatches 1; Indels 1; Gaps 1;

QY 2 QYIKANSKFI-CITE 15
Db 983 QYIKANSKFI-CITE 997

RESULT 4

MSPI_PLAF3 STANDARD; PRT; 1701 AA.
AC P13819;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE Merozoite surface protein 1 precursor (Merozoite surface antigens)
DE (PMMSA).
GN MSP-1.
OS Plasmodium falciparum (isolate FC27 / Papua New Guinea).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5837;
RN [1]

SEQUENCE FROM N.A.
RX MEDLINE=88142999; PubMed=2449612;
RA Peterson M.G., Coppel R.L., McIntyre P., Langford C.J., Woodrow G.,
RA Brown G.V., Anders R.F., Kemp D.J.;
RT "Variation in the precursor to the major merozoite surface antigens
RT of Plasmodium falciparum."
RL EMBO J. 6:4137-4142(1987).
RN [2]

Query Match 100.0%; Score 79; DB 1; Length 1314;
Best Local Similarity 100.0%; Pred. No. 2.7e-06;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MOYIKANSKFIGITEL 16
Db 828 MOYIKANSKFIGITEL 843

RESULT 2

VG84 BPML5 STANDARD; PRT; 66 AA.
AC Q05301; 1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 01-FEB-1994 (Rel. 28, Last annotation update)
DE Gene 84 protein (GP84).
GN 84.
OS Mycobacteriophage L5.
OC Viruses; dsDNA viruses, no RNA stage; Caudovirales; Siphoviridae;
OC L5-like viruses.
OX NCBI_TaxID=311757;
RN [1]

SEQUENCE FROM N.A.
RX MEDLINE=93211282; PubMed=8459766;
RA Hatfull G.F., Sarkis G.J.;
RT "DNA sequence, structure and gene expression of mycobacteriophage L5;
RL Mol. Microbiol. 7:395-405(1993).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; Z18946; CAA79460.1; -;
DR EMBL; S31029; S31029.
DR PIR; S31029; S31029.
SQ SEQUENCE 66 AA; 7424 MW; 9C7104C7A4FA74A5 CRC64;

Query Match 55.7%; Score 44; DB 1; Length 66;
Best Local Similarity 57.1%; Pred. No. 0.33;
Matches 8; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 3 YIKANSKFIGITEL 16
Db 50 YIKANSKFIGITEL 63

RESULT 3

MSPI_PLAF3 STANDARD; PRT; 1682 AA.
AC P19598; Q25921;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Merozoite surface protein 1 precursor (Merozoite surface antigens)
DE (PMMSA) (P190).
GN MSP-1.
OS Plasmodium falciparum (isolate ro-33 / Ghana).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5834;
RN [1]

SEQUENCE OF 1-1061 FROM N.A.
RX MEDLINE=88166657; PubMed=3327688;
RA Certa U., Rotmann D., Matile H., Reber-Liske R.;
RT "A naturally occurring gene encoding the major surface antigen
RT precursor p190 of Plasmodium falciparum lacks tripeptide repeats."
RL EMBO J. 6:4137-4142(1987).
RN [2]

```

RL  Eur. J. Biochem. 188:39-45(1990).
RN  [6]
RP  PARTIAL SEQUENCE.
RX  MEDLINE=92037649; PubMed=1935979;
RA  Krieglstein K.G., Henschen A.H., Weller U., Habermann E.;
RT  "Limited proteolysis of tetanus toxin. Relation to activity and
RT  identification of cleavage sites";
RL  Eur. J. Biochem. 202:41-51(1991).
RN  [7]
RP  IDENTIFICATION AS ZINC-PROTEASE.
RX  MEDLINE=93010948; PubMed=1396558;
RA  Schiavo G., Poulain B., Rossetto O., Benfenati F., Tauc L.,
RA  Montecucco C.;
RT  "Tetanus toxin is a zinc protein and its inhibition of
RT  neurotransmitter release and protease activity depend on zinc.";
RL  EMBO J. 11:3577-3583(1992).
RN  [8]
RP  IDENTIFICATION OF SUBSTRATE.
RX  MEDLINE=93063293; PubMed=1331807;
RA  Schiavo G., Benfenati F., Poulain B., Rossetto O., de Laureto P.P.,
RA  Dasgupta B.R., Montecucco C.;
RT  "Tetanus and botulinum-B neurotoxins block neurotransmitter release
RT  by proteolytic cleavage of synaptobrevin.";
RL  Nature 353:832-835(1992).
RN  [9]
RP  X-RAY CRYSTALLOGRAPHY (2.7 ANGSTROMS) OF 874-1314.
RX  MEDLINE=97475217; PubMed=9334741;
RA  Umland T.C., Wingert L.M., Swaminathan S., Furey W.F., Schmidt J.J.,
RA  Sax M.;
RT  "Structure of the receptor binding fragment HC of tetanus
RT  neurotoxin.";
RL  Nat. Struct. Biol. 4:788-792(1997).
CC  -!- FUNCTION: TETANUS TOXIN ACTS BY INHIBITING NEUROTRANSMITTER
CC  RELEASE. IT BINDS TO PERIPHERAL NEURONAL SYNAPSES, IS INTERNALIZED
CC  AND MOVES BY RETROGRADE TRANSPORT UP THE AXON INTO THE SPINAL CORD
CC  WHERE IT CAN MOVE BETWEEN POSTSYNAPTIC AND PRESYNAPTIC NEURONS. IT
CC  INHIBITS NEUROTRANSMITTER RELEASE BY ACTING AS A ZINC
CC  ENDOPEPTIDASE THAT CATALYZES THE HYDROLYSIS OF THE 76-GLN-|-PHE-77
CC  BOND OF SYNAPTOSREVIN-2.
CC  -!- CATALYTIC ACTIVITY: Hydrolysis of 76-Gln-|-Phe-77 bond in
CC  synaptobrevin 2.
CC  -!- COFACTOR: Binds 1 zinc ion per subunit (By similarity).
CC  -!- SUBUNIT: THE PRECURSOR POLYPEPTIDE IS SUBSEQUENTLY CLEAVED TO
CC  YIELD SUBCHAINS L AND H. THESE REMAIN LINKED BY A DISULFIDE BRIDGE
CC  AND ARE NON-TOXIC AFTER SEPARATION.
CC  -!- MISCELLANEOUS: THE C-TERMINAL OF THE HEAVY CHAIN BINDS TO
CC  GLYCOSIDE RECEPTORS.
CC  -!- SIMILARITY: Belongs to peptidase family M27.
CC  -----
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC  -----
DR  EMBL; X04436; CAA28033.1; -
DR  EMBL; X06214; CAA29564.1; -
DR  EMBL; AF528097; AAC37454.1; -
DR  EMBL; M12739; AAZ3282.1; -
DR  FIC; A25689; BTCLTN.
DR  PDB; 1AF9; 29-APR-98.
DR  PDB; 1A8D; 14-OCT-98.
DR  PDB; 1DQO; 27-MAR-00.
DR  PDB; 1DFO; 24-MAR-00.
DR  PDB; 1DIW; 24-MAR-00.
DR  PDB; 1DLL; 24-MAR-00.
DR  PDB; 1FV3; 05-SEP-01.
DR  MEROPS; M27.001; -
DR  InterPro; IPR008985; CoxA like lec_gl.
DR  InterPro; IPR002160; Kunitz legume.
DR  InterPro; IPR006025; Pept_M_Zn_BS.

```

```

DR  InterPro; IPR000395; Peptidase_M27.
DR  Pfam; PF01742; Peptidase_M27; 1.
DR  PRINTS; PR00760; BONTOKILYSIN.
DR  ProDom; PD001963; Bontoxilysin; 1.
DR  PROSITE; PS00142; ZINC_PROTEASE; 1.
KW  Neurotoxin, Transmembrane, Hydrolase, Metalloprotease, Zinc, Plasmid;
KW  3D-structure; Complete proteome.
FT  INIT_MET 0
FT  CHAIN 1 456
FT  CHAIN 457 1314
FT  METAL 232 232
FT  ACT_SITE 233 233
FT  METAL 236 236
FT  TRANSMEM 226 246
FT  TRANSMEM 689 689
FT  DISULFID 438 466
FT  DISULFID 1076 1092
FT  HELIX 876 882
FT  TURN 883 883
FT  STRAND 884 891
FT  TURN 892 892
FT  STRAND 894 897
FT  STRAND 904 907
FT  TURN 909 910
FT  STRAND 912 915
FT  STRAND 920 925
FT  TURN 928 929
FT  STRAND 932 935
FT  HELIX 938 940
FT  TURN 941 946
FT  STRAND 949 956
FT  HELIX 962 968
FT  TURN 969 970
FT  STRAND 972 977
FT  STRAND 980 981
FT  HELIX 983 985
FT  STRAND 987 995
FT  TURN 996 997
FT  STRAND 998 1004
FT  TURN 1006 1007
FT  STRAND 1010 1016
FT  STRAND 1020 1020
FT  TURN 1021 1022
FT  STRAND 1031 1037
FT  TURN 1039 1040
FT  STRAND 1042 1047
FT  TURN 1048 1049
FT  STRAND 1050 1056
FT  TURN 1058 1059
FT  STRAND 1058 1074
FT  TURN 1079 1080
FT  STRAND 1082 1091
FT  HELIX 1097 1105
FT  TURN 1106 1107
FT  STRAND 1112 1112
FT  STRAND 1114 1114
FT  TURN 1116 1117
FT  STRAND 1120 1120
FT  TURN 1122 1122
FT  TURN 1123 1124
FT  STRAND 1127 1131
FT  HELIX 1132 1134
FT  TURN 1135 1136
FT  STRAND 1137 1141
FT  TURN 1144 1145
FT  STRAND 1148 1152
FT  STRAND 1155 1158
FT  TURN 1159 1162
FT  STRAND 1163 1166
FT  STRAND 1173 1178
FT  TURN 1184 1185
FT  STRAND 1188 1188
FT  STRAND 1190 1190

```

TETANUS TOXIN LIGHT CHAIN.
 TETANUS TOXIN HEAVY CHAIN.
 ZINC (CATALYTIC) (BY SIMILARITY).
 BY SIMILARITY.
 ZINC (CATALYTIC) (BY SIMILARITY).
 POTENTIAL.
 POTENTIAL.
 INTERCHAIN.

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 10, 2004, 08:58:53 ; Search time 2.92607 Seconds
(without alignments)
284.724 Million cell updates/sec

Title: US-09-848-834A-2

Perfect score: 79

Sequence: 1 MQYIKANSKFIGITEL 16

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_42.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	79	100.0	1314	1	TETX_CLOTE
2	44	55.7	66	1	VG84_EPMU5
3	42.5	53.8	1682	1	MSPI_PLAF3
4	42.5	53.8	1701	1	MSPI_PLAF3
5	42.5	53.8	1701	1	MSPI_PLAF3
6	42.5	53.8	1726	1	MSPI_PLAF3
7	42.5	53.8	1726	1	MSPI_PLAF3
8	41	51.9	204	1	PYRC_SERMA
9	41	51.9	384	1	LE12_THETN
10	41	51.9	899	1	ACPD_HAEIN
11	40	50.6	194	1	ACPD_HAEIN
12	40	50.6	213	1	KAD_MYCCA
13	40	50.6	601	1	PEFI_LACLC
14	40	50.6	601	1	PEPF_LACLA
15	40	50.6	644	1	VHJ9_YEAST
16	39	49.4	445	1	GNT1_HUMAN
17	39	49.4	447	1	GNT1_MOUSE
18	39	49.4	447	1	GNT1_RABIT
19	39	49.4	447	1	GNT1_RAT
20	39	49.4	505	1	GPMI_MYCPU
21	39	49.4	548	1	2AD1_SCHPO
22	38	48.1	256	1	YD83_METJA
23	38	48.1	287	1	TRUB_AQUAE
24	38	48.1	572	1	HEMA_F13HT
25	38	48.1	627	1	2AD2_SCHPO
26	37	46.8	191	1	Y096_HAEIN
27	37	46.8	195	1	ACPD_VIBVU
28	37	46.8	231	1	PYRF_MYCPE
29	37	46.8	445	1	ARGA_VIBVU
30	37	46.8	451	1	MURD_BASBU
31	37	46.8	490	1	Y032_BORBU
32	37	46.8	510	1	G6PD_ASPNG
33	37	46.8	511	1	G6PD_EMENI

RESULT 1

ID	TETX_CLOTE	STANDARD;	PRT;	1314 AA.
DT	13-AUG-1987 (Rel. 05, Created)			
DT	13-AUG-1987 (Rel. 05, Last sequence update)			
DT	10-OCT-2003 (Rel. 42, Last annotation update)			
DE	Tetanus toxin precursor (EC 3.4.24.68) (tencoxylisin) [Contains:			
DE	Tetanus toxin light chain (Tetanus toxin chain L); Tetanus toxin heavy			
DE	chain (Tetanus toxin chain H)]			
GN	TETX OR CTP60.			
OS	Clostridium tetani.			
OG	Plasmid pE88, and Plasmid 75 Kbp.			
OC	Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;			
OC	Clostridium.			
OX	NCBI_TaxID=1513;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	PLASMID=75 Kbp;			
RX	MEDLINE=87053814; PubMed=3536478;			
RA	Bisel U., Jarausch W., Goretzki K., Henschen A., Engels J.,			
RA	Weller U., Hudel M., Habermann E., Niemann H.;			
RT	"Tetanus toxin: primary structure, expression in E. coli, and			
RT	homology with botulinum toxins."			
RL	EMBO J. 5:2495-2502(1986).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=CN3911; PLASMID=75 Kbp;			
RX	MEDLINE=87040747; PubMed=374547;			
RA	Fairweather N.F., Lyness V.A.;			
RA	"The complete nucleotide sequence of tetanus toxin."			
RT	Nucleic Acids Res. 14:7809-7812(1986).			
RN	[3]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=Massachusetts / E88; PLASMID=pE88;			
RX	MEDLINE=22457253; PubMed=2552129;			
RA	Brueggemann H., Baumer S., Fricke W.F., Wietzer A., Liesegang H.,			
RA	Decker I., Herzberg C., Martinez-Arias R., Merkl R., Henne A.,			
RA	Gottschalk G.;			
RT	"The genome sequence of Clostridium tetani, the causative agent of			
RT	tetanus disease."			
RL	Proc. Natl. Acad. Sci. U.S.A. 100:1316-1321(2003).			
RN	[4]			
RP	SEQUENCE OF 742-1314 FROM N.A.			
RC	PLASMID=75 Kbp;			
RX	MEDLINE=86085672; PubMed=3510187;			
RA	Fairweather N.F., Lyness V.A., Pickard D.J., Allen G., Thomson R.O.;			
RT	"Cloning, nucleotide sequencing, and expression of tetanus toxin			
RT	fragment C in Escherichia coli."			
RL	J. Bacteriol. 165:21-27(1986).			
RN	[5]			
RP	PARTIAL SEQUENCE, AND DISULFIDE BONDS.			
RC	MEDLINE=90201034; PubMed=2108021;			
RA	Kriegelstein K., Henschen A., Weller U., Habermann E.;			
RT	"Arrangement of disulfide bridges and positions of sulphydryl groups			
RT	in tetanus toxin."			

ALIGNMENTS

34	37	46.8	757	1	2A5D_YEAST
35	37	46.8	774	1	RBP3_INCB
36	37	46.8	774	1	RBP3_INCB
37	36	45.6	169	1	Y358_BUCAL
38	36	45.6	247	1	C561_XENLA
39	36	45.6	258	1	MIP_CHLPN
40	36	45.6	296	1	YD01_CLOAB
41	36	45.6	329	1	NADA_CVAPA
42	36	45.6	333	1	DPOB_XENLA
43	36	45.6	461	1	NIFN_RHOCA
44	36	45.6	495	1	G6PD_PICUA
45	36	45.6	630	1	YND1_YEAST

P33903	saccharomyc
P21770	influenza c
P13877	influenza c
P57439	buchnera ap
Q91577	xenopus lae
Q977p3	chlamydia p
P33659	clostridium
P31179	cyanophora
O57383	xenopus lae
P19077	rhodobacter
P11410	pichia jadi
P40009	saccharomyc

Wed Mar 10 10:34:30 2004

us-09-848-834a-2.open.rpr

Oy 3 YIKANSKFIGITEL 16
|||
Db 57 YIKYNEKFGITTL 70

Search completed: March 10, 2004, 09:16:40
Job time : 7.03937 secs

A:Residues: 1-1086 <OLA>
A:Cross-references: EMBL:X61930
A:Experimental source: isolate RO-71
C:Genetics:
A:Gene: MSAL
C:Superfamily: major merozoite surface antigen
C:Keywords: glycoprotein; merozoite; surface antigen; tandem repeat

Query Match 53.8%; Score 42.5; DB 2; Length 1086;
Best Local Similarity 60.0%; Pred.No.26;
Matches 9; Conservative 4; Mismatches 1; Indels 1; Gaps 1;

QY 2 QYKANSKFI-GITE 15
|:|:|:|:|:|:|
Db 1009 QFVKSNSKVITGLTE 1023

RESULT 9
A54498
major merozoite surface antigen precursor - malaria parasite (Plasmodium falciparum) (C)
C:Species: Plasmodium falciparum
C:Date: 28-Oct-1994 #sequence_revision 28-Oct-1994 #text_change 09-Jun-2000
C:Accession: A54498
R:Petersen, M.G.; Coppel, R.L.; McIntyre, P.; Langford, C.J.; Woodrow, G.; Brown, G.V.
Mol. Biochem. Parasitol. 27, 291-302, 1988
A>Title: Variation in the precursor to the major merozoite surface antigens of Plasmodium
A:Reference number: A54498; MUID:86142999; PMID:2449612
A:Accession: A54498
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1701 <PET>
A:Cross-references: GB:SI610413; PIDN:AAA29653.1; PID:gi60413
C:Superfamily: major merozoite surface antigen
C:Keywords: surface antigen

Query Match 53.8%; Score 42.5; DB 2; Length 1701;
Best Local Similarity 60.0%; Pred.No.41;
Matches 9; Conservative 4; Mismatches 1; Indels 1; Gaps 1;

QY 2 QYKANSKFI-GITE 15
|:|:|:|:|:|:|
Db 1001 QFVKSNSKVITGLTE 1015

RESULT 10
A26868
major merozoite surface antigen precursor - malaria parasite (Plasmodium falciparum) (C)
C:Species: Plasmodium falciparum
C:Date: 19-Nov-1988 #sequence_revision 19-Nov-1988 #text_change 09-Jun-2000
C:Accession: A26868
R:Tanabe K.; Mackay, M.; Goman, M.; Scaife, J.G.
J. Mol. Biol. 195, 273-287, 1987
A>Title: Allelic dimorphism in a surface antigen gene of the malaria parasite Plasmodium
A:Reference number: A26868; MUID:88011243; PMID:3079521
A:Accession: A26868
A:Molecule type: DNA
A:Residues: 1-1701 <TAN>
C:Superfamily: major merozoite surface antigen
C:Keywords: surface antigen
F:1-19/DNA: signal sequence #status predicted <SIG>
F:20-1701/Product: major merozoite surface antigen #status predicted <MAT>

Query Match 53.8%; Score 42.5; DB 2; Length 1701;
Best Local Similarity 60.0%; Pred.No.41;
Matches 9; Conservative 4; Mismatches 1; Indels 1; Gaps 1;

QY 2 QYKANSKFI-GITE 15
|:|:~|:|:|:|:|:|
Db 1001 QFVKSNSKVITGLTE 1015

RESULT 11
SAZQGM

R;de Filippis, V.; Vangelista, L.; Schiavo, G.; Tonello, F.; Montecucco, C.
 Eur. J. Biochem. 229, 61-69, 1995
 A;Title: Structural studies on the zinc-endopeptidase light chain of tetanus neurotoxin.
 A;Reference number: S69348; MUID:95262688; PMID:7744050
 A;Accession: S69348
 A;Molecule type: protein
 A;Residues: 2-31 <DEF>
 C;Comment: The source of this protein was an extrachromosomal plasmid.
 C;Comment: The precursor is cleaved by endogenous proteinase activity to form light (fra
 dual chains are not toxic when separated). The amino end of the heavy chain (fragment B)
 C;Comment: Fragment B forms ion channels in a lipid bilayer. Fragment C binds to ganglio
 C;Comment: This potent neurotoxin binds to peripheral neuronal synapses, is internalized
 presynaptic neurons. It inhibits neurotransmitter release by proteolytic cleavage of sy
 C;Function:
 C;Description: blocks neuroexocytosis via hydrolysis of a Gln-Phe peptide bond in synap
 C;Superfamily: tetanus toxin
 C;Keywords: hydrolase; metalloproteinase; neurotoxin; transmembrane protein; zinc
 F;2-457/Product: tetoxylysin light chain (fragment A) #status predicted <TT>
 F;461-1315/Product: tetoxylysin heavy chain (fragment B.C) #status experimental <TTH>
 F;461-864/Domain: channel forming (fragment B) #status predicted <TXB>
 F;865-1315/Domain: ganglioside binding (fragment C) #status predicted <TXC>
 F;232,237/Binding site: zinc (His) #status predicted
 F;234/Active site: Glu #status predicted

Query Match 100.0%; Score 79; DB 1; Length 1315;
 Best Local Similarity 100.0%; Pred. No. 8.5e-06;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MOYIKANSKFIGITEL 16
 |||||
 DB 829 MOYIKANSKFIGITEL 844
 |||||

RESULT 2
 S29982
 class II histocompatibility antigen - Atlantic salmon
 C;Species: Salmo salar (Atlantic salmon)
 C;Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 20-Sep-1999
 C;Accession: S29982
 R;Hordvik, I.
 submitted to the EMBL Data Library, October 1992
 A;Reference number: S29980
 A;Accession: S29982
 A;Status: preliminary
 A;Molecule type: mRNA
 A;Residues: 1-244 <HOR>
 A;Cross-references: EMBL:X70156; NID:g64369; PID:g64370
 C;Superfamily: class II histocompatibility antigen; immunoglobulin homology

Query Match 56.3%; Score 44.5; DB 2; Length 244;
 Best Local Similarity 55.6%; Pred. No. 2.6;
 Matches 10; Conservative 3; Mismatches 2; Indels 3; Gaps 1;

QY 2 QYIKANS--KFIGITEL 16
 :||: |||||
 DB 51 EYIRFNSVTGKFGVGYTEL 68.
 :||: |||||

RESULT 3
 S31029
 gene 84 protein - Mycobacterium phage L5
 C;Species: Mycobacterium phage L5
 C;Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 08-Oct-1999
 C;Accession: S31029
 R;Donnelly-Wu, M.K.; Jacobs Jr., W.R.; Hatfull, G.F.
 Mol. Microbiol. 7, 407-417, 1993
 A;Title: Superinfection immunity of mycobacteriophage L5: applications for genetic trans
 A;Reference number: S30949; MUID:93211283; PMID:8459767
 A;Accession: S31029
 A;Status: nucleic acid sequence not shown; translation not shown
 A;Molecule type: DNA
 A;Residues: 1-66 <DON>
 A;Cross-references: EMBL:Z18946; NID:g15859; PIDN:CAA79460.1; PID:e59702; PID:g579152

A;Note: the nucleotide sequence was submitted to the EMBL Data Library, December 1992
 C;Genetics:
 A;Gene: 84
 A;Start codon: GTG

Query Match 55.7%; Score 44; DB 2; Length 66;
 Best Local Similarity 57.1%; Pred. No. 0.86;
 Matches 8; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 3 YIKANSKFIGITEL 16
 |||||
 DB 50 YIKENKFKVGTWEV 63
 |||||

RESULT 4

G86826
 diamine N-acetyltransferase (EC 2.3.1.57) [imported] - Lactococcus lactis subsp. lacti
 N;Alternate names: spermidine acetyltransferase
 C;Species: Lactococcus lactis subsp. lactis
 C;Date: 23-Mar-2001 #sequence_revision 23-Mar-2001 #text_change 17-May-2002
 C;Accession: G86826
 R;Bolotin, A.; Wincker, P.; Mauger, S.; Jaillon, O.; Malarme, K.; Weissenbach, J.; Ehr:
 Genoms Res. 11, 731-753, 2001
 A;Title: The complete genome sequence of the lactic acid bacterium Lactococcus lactis
 A;Reference number: A86625; MUID:21235186; PMID:11337471
 A;Accession: G86826
 A;Status: preliminary
 A;Molecule type: DNA
 A;Residues: 1-190 <STO>
 A;Cross-references: GB:AB005176; PID:gl2724622; PIDN:AAK05713.1; GSPDB:GN00146
 A;Experimental source: strain IL1403
 C;Genetics:
 A;Gene: ygfF
 C;Superfamily: Escherichia coli ribosomal-protein-alanine N-acetyltransferase rimJ
 C;Keywords: acyltransferase; coenzyme A

Query Match 54.4%; Score 43; DB 2; Length 180;
 Best Local Similarity 69.2%; Pred. No. 3.6;
 Matches 9; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 4 IKANSKFIGITEL 16
 :||: |||||
 DB 65 IEANDTFIGIVEL 77
 :||: |||||

RESULT 5

T42976
 hypothetical protein 63 - ateline herpesvirus 3 (strain 73)
 C;Species: ateline herpesvirus 3
 A;Variety: strain 73
 C;Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 21-Jan-2000
 C;Accession: T42976
 R;Albrecht, J.C.; Fleckenstein, B.
 submitted to the EMBL Data Library, August 1998
 A;Description: Primary structure of the herpesvirus ateles genome.
 A;Reference number: Z22274
 A;Accession: T42976
 A;Status: preliminary; translated from GB/EMBL/DBJ
 A;Molecule type: DNA
 A;Residues: 1-899 <ALB>
 A;Cross-references: EMBL:AF083424; PIDN:AAC95587.1
 A;Experimental source: strain 73

Query Match 54.4%; Score 43; DB 2; Length 899;
 Best Local Similarity 64.3%; Pred. No. 18;
 Matches 9; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 2 QYIKANSKFIGITE 15
 |||||
 DB 124 QYITNSFTGTOTE 137
 |||||

RESULT 6

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 10, 2004, 08:58:54 ; Search time 4.98054 Seconds
(without alignments)
309.015 Million cell updates/sec

Title: US-09-848-834A-2

Perfect score: 79

Sequence: 1 MQYIKANSRFIGITEL 16

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: Pir1.*

2: Pir2.*

3: Pir3.*

4: Pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Length	DB ID	Description
1	79	100.0	1315	1 BFLCTN
2	44.5	56.3	244	2 S29982
3	44	55.7	66	3 S31029
4	43	54.4	180	4 G86826
5	43	54.4	899	5 T42976
6	42.5	53.8	82	6 I50966
7	42.5	53.8	1060	7 S06286
8	42.5	53.8	1086	8 S16752
9	42.5	53.8	1701	9 A54458
10	42.5	53.8	1701	10 A26828
11	42.5	53.8	1726	11 S4200M
12	42.5	53.8	1726	12 A45948
13	42	53.2	1333	13 S38635
14	41	51.9	79	14 D58794
15	41	51.9	123	15 G48677
16	41	51.9	447	16 H97146
17	41	51.9	899	17 G36812
18	40.5	51.3	245	18 S29980
19	40	50.6	194	19 G44026
20	40	50.6	213	20 K1YMC
21	40	50.6	501	21 A86158
22	40	50.6	501	22 T52135
23	40	50.6	601	23 A55485
24	40	50.6	601	24 G86840
25	40	50.6	644	25 S46746
26	39	49.4	102	26 P41491
27	39	49.4	119	27 PH1516
28	39	49.4	119	28 PH1518
29	39	49.4	119	29 PH1519

30 39 49.4 123 2 F48677
31 39 49.4 135 2 PH1494
32 39 49.4 140 2 PH1488
33 39 49.4 189 2 G97978
34 39 49.4 225 2 A82539
35 39 49.4 226 2 B71808
36 39 49.4 349 2 T43043
37 39 49.4 381 2 F71196
38 39 49.4 383 2 F51466
39 39 49.4 423 2 F64690
40 39 49.4 445 1 XUHUMB
41 39 49.4 447 1 A42500
42 39 49.4 447 1 A38561
43 39 49.4 447 2 JC2076
44 39 49.4 492 2 AH2079
45 39 49.4 505 2 C90569

ig heavy chain V-D
ig heavy chain V r
ig heavy chain V r
conserved hypothet
hypothetical prote
type II restrictio
probable acetyl-Co
probable hexosyltr
hypothetical prote
type IIS restricti
alpha-1,3-mannosyl
alpha-1,3-mannosyl
alpha-1,3-mannosyl
alpha-1,3-mannosyl
alpha-1,3-mannosyl
alpha-1,3-mannosyl
hypothetical prote

precursor - Clostridium tetani
tetanospasmin toxin
N:Alternate names: tetanus neurotoxin
C:Species: Clostridium tetani
C:Date: 31-Mar-1988 #sequence revision 31-Mar-1988 #text change 03-Jun-2002
C:Accession: A25689; A25757; A25194; A60759; S69348; S09364
R:Eisel, U.; Jarausch, W.; Gorsetski, K.; Henschen, A.; Engels, J.; Weller, U.; Hudel, I.
EMBO J. 5, 2495-2502, 1986
A:Title: Tetanus toxin: Primary structure, expression in E. coli, and homology with bo.
A:Reference number: A25689; MUID:87053814; PMID:3536478
A:Accession: A25689
A:Molecule type: DNA
A:Residues: 1-1315 <EIS>
A:Cross-references: GB:X04436; NID:940769; PIDN:CAA28033.1; PID:940770
R:Fairweather, N.F.; Lyness, V.A.
Nucleic Acids Res. 14, 7809-7812, 1986
A:Title: The complete nucleotide sequence of tetanus toxin.
A:Reference number: A25757; MUID:87040747; PMID:3774547
A:Accession: A25757
A:Molecule type: DNA
A:Residues: 1-1315 <FAI>
A:Cross-references: GB:X06214; NID:940773; PIDN:CAA29564.1; PID:940774
A:Experimental source: strain CN3911
R:Fairweather, N.F.; Lyness, V.A.; Pickard, D.J.; Allen, G.; Thomson, R.O.
J. Bacteriol. 165, 21-27, 1986
A:Title: Cloning, nucleotide sequencing, and expression of tetanus toxin fragment C in
A:Reference number: A25194; MUID:86085672; PMID:3510187
A:Accession: A25194
A:Molecule type: DNA
A:Residues: 743-1315 <FA2>
A:Cross-references: GB:M12739; NID:9144920; PIDN:AAA23282.1; PID:9144921
A:Accession: B25194
A:Molecule type: protein
A:Residues: 865-894 <FA3>
R:Natsuda, M.; Lei, D.B.; Sugimoto, N.; Ozutsumi, K.; Okabe, T.
Infect. Immun. 57, 3588-3593, 1989
A:Title: Isolation, purification, and characterization of fragment B, the NH-2-termina.
A:Reference number: A60759; MUID:90035436; PMID:2478476
A:Accession: A60759
A:Molecule type: protein
A:Residues: 461-475 <WAT>
R:Demetz, S.; Lanzavecchia, L.; Eisel, U.; Niemann, H.; Widmann, C.; Corradin, G.
J. Immunol. 142, 394-402, 1989
A:Title: Delination of several DR-restricted tetanus toxin T cell epitopes.
A:Reference number: JS0098; MUID:89093918; PMID:2463305
A:Contents: annotation; epitope region
R:Schiavo, G.; Benfenati, F.; Poulain, B.; Rossetto, O.; de Laureto, P.P.; DasGupta, B.
Nature 359, 832-835, 1992
A:Title: Tetanus and botulinum-B neurotoxins block neurotransmitter release by proteoly.
A:Reference number: S27125; MUID:93063293; PMID:1331807
A:Contents: annotation

ALIGNMENTS

RESULT 1

BTCLTN
tetanospasmin (EC 3.4.24.68) precursor - Clostridium tetani
N:Alternate names: tetanus neurotoxin
C:Species: Clostridium tetani
C:Date: 31-Mar-1988 #sequence revision 31-Mar-1988 #text change 03-Jun-2002
C:Accession: A25689; A25757; A25194; A60759; S69348; S09364
R:Eisel, U.; Jarausch, W.; Gorsetski, K.; Henschen, A.; Engels, J.; Weller, U.; Hudel, I.
EMBO J. 5, 2495-2502, 1986
A:Title: Tetanus toxin: Primary structure, expression in E. coli, and homology with bo.
A:Reference number: A25689; MUID:87053814; PMID:3536478
A:Accession: A25689
A:Molecule type: DNA
A:Residues: 1-1315 <EIS>
A:Cross-references: GB:X04436; NID:940769; PIDN:CAA28033.1; PID:940770
R:Fairweather, N.F.; Lyness, V.A.
Nucleic Acids Res. 14, 7809-7812, 1986
A:Title: The complete nucleotide sequence of tetanus toxin.
A:Reference number: A25757; MUID:87040747; PMID:3774547
A:Accession: A25757
A:Molecule type: DNA
A:Residues: 1-1315 <FAI>
A:Cross-references: GB:X06214; NID:940773; PIDN:CAA29564.1; PID:940774
A:Experimental source: strain CN3911
R:Fairweather, N.F.; Lyness, V.A.; Pickard, D.J.; Allen, G.; Thomson, R.O.
J. Bacteriol. 165, 21-27, 1986
A:Title: Cloning, nucleotide sequencing, and expression of tetanus toxin fragment C in
A:Reference number: A25194; MUID:86085672; PMID:3510187
A:Accession: A25194
A:Molecule type: DNA
A:Residues: 743-1315 <FA2>
A:Cross-references: GB:M12739; NID:9144920; PIDN:AAA23282.1; PID:9144921
A:Accession: B25194
A:Molecule type: protein
A:Residues: 865-894 <FA3>
R:Natsuda, M.; Lei, D.B.; Sugimoto, N.; Ozutsumi, K.; Okabe, T.
Infect. Immun. 57, 3588-3593, 1989
A:Title: Isolation, purification, and characterization of fragment B, the NH-2-termina.
A:Reference number: A60759; MUID:90035436; PMID:2478476
A:Accession: A60759
A:Molecule type: protein
A:Residues: 461-475 <WAT>
R:Demetz, S.; Lanzavecchia, L.; Eisel, U.; Niemann, H.; Widmann, C.; Corradin, G.
J. Immunol. 142, 394-402, 1989
A:Title: Delination of several DR-restricted tetanus toxin T cell epitopes.
A:Reference number: JS0098; MUID:89093918; PMID:2463305
A:Contents: annotation; epitope region
R:Schiavo, G.; Benfenati, F.; Poulain, B.; Rossetto, O.; de Laureto, P.P.; DasGupta, B.
Nature 359, 832-835, 1992
A:Title: Tetanus and botulinum-B neurotoxins block neurotransmitter release by proteoly.
A:Reference number: S27125; MUID:93063293; PMID:1331807
A:Contents: annotation

mismatches	8:	Conservative	3:	Mismatches	3:	Indels	0:	Gaps	0:
------------	----	--------------	----	------------	----	--------	----	------	----

RA Adachi J., Aizawa K., Akahira S., Akimura T., Arai A., Aono H.,
RA Arakawa T., Bono H., Carninci P., Fukuda S., Fukunishi Y., Furuno M.,
RA Hanagaki T., Hara A., Hayatsu N., Hiramoto K., Hiraoka T., Hori F.,
RA Imorani K., Ishii Y., Itoh M., Izawa M., Kasukawa T., Kato H.,
RA Kawai J., Kojima Y., Konno H., Kouda M., Koya S., Kurihara C.,
RA Matsuyama T., Miyazaki A., Nihi K., Nemura K., Numazaki R., Ohno M.,
RA Okazaki Y., Okido T., Owa C., Saito H., Saito R., Sakai C., Sakai K.,
RA Sano H., Sasaki D., Shibata K., Shibata Y., Shingawa A., Shiraki T.,
RA Sogabe Y., Suzuki H., Tagami M., Tagawa A., Takahashi F., Tanaka T.,
RA Teijima Y., Toyota T., Yamamura T., Yasunishi A., Yoshida K., Yoshino M.,
RA Muramatsu M., Hayashizaki Y.;
RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
RL [2]
RL [3]
RL [4]
RL [5]
RL [6]
RL [7]
RL [8]
RL [9]
RL [10]
RL [11]
RL [12]
RL [13]
RL [14]
RL [15]
RL [16]
RL [17]
RL [18]
RL [19]
RL [20]
RL [21]
RL [22]
RL [23]
RL [24]
RL [25]
RL [26]
RL [27]
RL [28]
RL [29]
RL [30]
RL [31]
RL [32]
RL [33]
RL [34]
RL [35]
RL [36]
RL [37]
RL [38]
RL [39]
RL [40]
RL [41]
RL [42]
RL [43]
RL [44]
RL [45]
RL [46]
RL [47]
RL [48]
RL [49]
RL [50]
RL [51]
RL [52]
RL [53]
RL [54]
RL [55]
RL [56]
RL [57]
RL [58]
RL [59]
RL [60]
RL [61]
RL [62]
RL [63]
RL [64]
RL [65]
RL [66]
RL [67]
RL [68]
RL [69]
RL [70]
RL [71]
RL [72]
RL [73]
RL [74]
RL [75]
RL [76]
RL [77]
RL [78]
RL [79]
RL [80]
RL [81]
RL [82]
RL [83]
RL [84]
RL [85]
RL [86]
RL [87]
RL [88]
RL [89]
RL [90]
RL [91]
RL [92]
RL [93]
RL [94]
RL [95]
RL [96]
RL [97]
RL [98]
RL [99]
RL [100]
RL [101]
RL [102]
RL [103]
RL [104]
RL [105]
RL [106]
RL [107]
RL [108]
RL [109]
RL [110]
RL [111]
RL [112]
RL [113]
RL [114]
RL [115]
RL [116]
RL [117]
RL [118]
RL [119]
RL [120]
RL [121]
RL [122]
RL [123]
RL [124]
RL [125]
RL [126]
RL [127]
RL [128]
RL [129]
RL [130]
RL [131]
RL [132]
RL [133]
RL [134]
RL [135]
RL [136]
RL [137]
RL [138]
RL [139]
RL [140]
RL [141]
RL [142]
RL [143]
RL [144]
RL [145]
RL [146]
RL [147]
RL [148]
RL [149]
RL [150]
RL [151]
RL [152]
RL [153]
RL [154]
RL [155]
RL [156]
RL [157]
RL [158]
RL [159]
RL [160]
RL [161]
RL [162]
RL [163]
RL [164]
RL [165]
RL [166]
RL [167]
RL [168]
RL [169]
RL [170]
RL [171]
RL [172]
RL [173]
RL [174]
RL [175]
RL [176]
RL [177]
RL [178]
RL [179]
RL [180]
RL [181]
RL [182]
RL [183]
RL [184]
RL [185]
RL [186]
RL [187]
RL [188]
RL [189]
RL [190]
RL [191]
RL [192]
RL [193]
RL [194]
RL [195]
RL [196]
RL [197]
RL [198]
RL [199]
RL [200]
RL [201]
RL [202]
RL [203]
RL [204]
RL [205]
RL [206]
RL [207]
RL [208]
RL [209]
RL [210]
RL [211]
RL [212]
RL [213]
RL [214]
RL [215]
RL [216]
RL [217]
RL [218]
RL [219]
RL [220]
RL [221]
RL [222]
RL [223]
RL [224]
RL [225]
RL [226]
RL [227]
RL [228]
RL [229]
RL [230]
RL [231]
RL [232]
RL [233]
RL [234]
RL [235]
RL [236]
RL [237]
RL [238]
RL [239]
RL [240]
RL [241]
RL [242]
RL [243]
RL [244]
RL [245]
RL [246]
RL [247]
RL [248]
RL [249]
RL [250]
RL [251]
RL [252]
RL [253]
RL [254]
RL [255]
RL [256]
RL [257]
RL [258]
RL [259]
RL [260]
RL [261]
RL [262]
RL [263]
RL [264]
RL [265]
RL [266]
RL [267]
RL [268]
RL [269]
RL [270]
RL [271]
RL [272]
RL [273]
RL [274]
RL [275]
RL [276]
RL [277]
RL [278]
RL [279]
RL [280]
RL [281]
RL [282]
RL [283]
RL [284]
RL [285]
RL [286]
RL [287]
RL [288]
RL [289]
RL [290]
RL [291]
RL [292]
RL [293]
RL [294]
RL [295]
RL [296]
RL [297]
RL [298]
RL [299]
RL [300]
RL [301]
RL [302]
RL [303]
RL [304]
RL [305]
RL [306]
RL [307]
RL [308]
RL [309]
RL [310]
RL [311]
RL [312]
RL [313]
RL [314]
RL [315]
RL [316]
RL [317]
RL [318]
RL [319]
RL [320]
RL [321]
RL [322]
RL [323]
RL [324]
RL [325]
RL [326]
RL [327]
RL [328]
RL [329]
RL [330]
RL [331]
RL [332]
RL [333]
RL [334]
RL [335]
RL [336]
RL [337]
RL [338]
RL [339]
RL [340]
RL [341]
RL [342]
RL [343]
RL [344]
RL [345]
RL [346]
RL [347]
RL [348]
RL [349]
RL [350]
RL [351]
RL [352]
RL [353]
RL [354]
RL [355]
RL [356]
RL [357]
RL [358]
RL [359]
RL [360]
RL [361]
RL [362]
RL [363]
RL [364]
RL [365]
RL [366]
RL [367]
RL [368]
RL [369]
RL [370]
RL [371]
RL [372]
RL [373]
RL [374]
RL [375]
RL [376]
RL [377]
RL [378]
RL [379]
RL [380]
RL [381]
RL [382]
RL [383]
RL [384]
RL [385]
RL [386]
RL [387]
RL [388]
RL [389]
RL [390]
RL [391]
RL [392]
RL [393]
RL [394]
RL [395]
RL [396]
RL [397]
RL [398]
RL [399]
RL [400]
RL [401]
RL [402]
RL [403]
RL [404]
RL [405]
RL [406]
RL [407]
RL [408]
RL [409]
RL [410]
RL [411]
RL [412]
RL [413]
RL [414]
RL [415]
RL [416]
RL [417]
RL [418]
RL [419]
RL [420]
RL [421]
RL [422]
RL [423]
RL [424]
RL [425]
RL [426]
RL [427]
RL [428]
RL [429]
RL [430]
RL [431]
RL [432]
RL [433]
RL [434]
RL [435]
RL [436]
RL [437]
RL [438]
RL [439]
RL [440]
RL [441]
RL [442]
RL [443]
RL [444]
RL [445]
RL [446]
RL [447]
RL [448]
RL [449]
RL [450]
RL [451]
RL [452]
RL [453]
RL [454]
RL [455]
RL [456]
RL [457]
RL [458]
RL [459]
RL [460]
RL [461]
RL [462]
RL [463]
RL [464]
RL [465]
RL [466]
RL [467]
RL [468]
RL [469]
RL [470]
RL [471]
RL [472]
RL [473]
RL [474]
RL [475]
RL [476]
RL [477]
RL [478]
RL [479]
RL [480]
RL [481]
RL [482]
RL [483

Query Match 54.4%; Score 43; DB 11; Length 291;
Best Local Similarity 57.1%; Pred. No. 25;

Matches 9; Conservative 4; Mismatches 2; Indels 3; Gaps 1;

Qy 2 QYKANS---KFIGITEL 16
: : : : :
Db 16 EYVRFNSTVGKFGVGYTEL 33

RESULT 12

Q9CF66 PRELIMINARY; PRT; 180 AA.
ID Q9CF66
AC Q9CF66
DT 01-JUN-2001 (TREMELrel. 17, Created)
DT 01-JUN-2001 (TREMELrel. 17, Last sequence update)
DT 01-OCT-2003 (TREMELrel. 25, Last annotation update)
DE Spermidine acetyltransferase (EC 2.3.1.57).
GN YQFF OR L1615.
OS Lactococcus lactis (subsp. lactis) (Streptococcus lactis).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae; Lactococcus.
OX NCBI_TaxID=1360;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=IL1403;
RX MEDLINE=21235186; PubMed=11337471;
RA Boletín A., Wincker P., Mauger S., Jaillon O., Malarne K.,
FA Weissbach J., Ehrlich S.D., Sorokin A.;
RT "The complete genome sequence of the lactic acid bacterium Lactococcus
lactis ssp. lactis IL1403."
RL Genome Res. 11:731-753(2001).
DR EMBL; AE006391; AAK05713.1; -
DR PIR; G86826; G86826.
DR GO; GO:0004145; F:diamine N-acetyltransferase activity; IEA.
DR GO; GO:0016740; F:transferase activity; IEA.
DR InterPro; IPR000182; GCSacetyl_trans.
DR Pfam; PF00583; Acetyltransf; 1.
KW Transferase; Complete proteome.
SQ SEQUENCE 180 AA; 21022 MW; 6BBD148524C0DF3C CRC64;

Query Match 54.4%; Score 43; DB 16; Length 180;
Best Local Similarity 69.2%; Pred. No. 15;
Matches 9; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy 4 IKANSKFIGITEL 16
: : : : :
Db 65 IEANDTFIGIVEL 77

RESULT 13

Q9MCL7 PRELIMINARY; PRT; 250 AA.
ID Q9MCL7
AC Q9MCL7
DT 01-OCT-2000 (TREMELrel. 15, Created)
DT 01-OCT-2000 (TREMELrel. 15, Last sequence update)
DT 01-OCT-2003 (TREMELrel. 25, Last annotation update)
DE ORF13.
GN ORF13.
OS Streptococcus thermophilus bacteriophage 7201.
OC Viruses; dsDNA viruses, no RNA stage; Caudovirales; Siphoviridae.
OX NCBI_TaxID=112023;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20088830; PubMed=16206678;
RA Stanley E., Walsh L., van der Zwet A., Fitzgerald G.F.,
RA van Sinderen D.;
RT "Identification of four loci isolated from two Streptococcus
thermophilus phage genomes responsible for mediating bacteriophage
resistance";
RL FEMS Microbiol. Lett. 182:271-277(2000).
RN [2]
RP SEQUENCE FROM N.A.
RA Stanley E., Walsh L., Fitzgerald G.F., van Sinderen D.;
RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF145054; RAF43506.1; -
DR GO; GO:0003677; F:DNA binding; IEA.

DR GO; GO:0008170; F:N-methyltransferase activity; IEA.

DR GO; GO:0006306; P:DNA methylation; IEA.
DR InterPro; IPR002295; D2IN6_mtfase.
DR InterPro; IPR001091; Met_trans_CN4.
DR Pfam; PF01555; N6/N4_Mtase.
DR PRINTS; PR00506; D2IN6MTFRASE.
DR PRINTS; PR00508; S2IN4MTFRASE.
SQ SEQUENCE 250 AA; 28744 MW; 725EA3B2B56B0D7F CRC64;

Query Match 54.4%; Score 43; DB 9; Length 250;
Best Local Similarity 57.1%; Pred. No. 21;
Matches 8; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

Qy 2 QYKANSKFIGITE 15
: : : : :
Db 131 QVILKANKIVGATE 144

RESULT 14

Q9XJEB PRELIMINARY; PRT; 252 AA.
ID Q9XJEB
AC Q9XJEB;
DT 01-NOV-1999 (TREMELrel. 12, Created)
DT 01-NOV-1999 (TREMELrel. 12, Last sequence update)
DT 01-OCT-2003 (TREMELrel. 25, Last annotation update)
DE Putative methylase.
OS Lactococcus lactis bacteriophage Tuc2009.
OC Viruses; dsDNA viruses, no RNA stage; Caudovirales; Siphoviridae.
OX NCBI_TaxID=35241;
RN [1]
RP SEQUENCE FROM N.A.
RA van Sinderen D., van de Guchte M., Seegers J.F.M.L., Fitzgerald G.F.;
RT "Molecular analysis of the temperate lactococcal phage Tuc2009."
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF109874; AAD37103.1; -
DR GO; GO:0003677; F:DNA binding; IEA.
DR GO; GO:0008170; F:N-methyltransferase activity; IEA.
DR GO; GO:0006306; P:DNA methylation; IEA.
DR InterPro; IPR002295; D2IN6_mtfase.
DR InterPro; IPR001091; Met_trans_CN4.
DR Pfam; PF01555; N6/N4_Mtase.
DR PRINTS; PR00506; D2IN6MTFRASE.
DR PRINTS; PR00508; S2IN4MTFRASE.
KW Methyltransferase.
SQ SEQUENCE 252 AA; 29357 MW; 4D5A59FF47363948 CRC64;

Query Match 54.4%; Score 43; DB 9; Length 252;
Best Local Similarity 57.1%; Pred. No. 22;
Matches 8; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

Qy 2 QYKANSKFIGITE 15
: : : : :
Db 130 QVILKANKIVGATE 143

RESULT 15

Q9CRV4 PRELIMINARY; PRT; 291 AA.
ID Q9CRV4
AC Q9CRV4;
DT 01-JUN-2001 (TREMELrel. 17, Created)
DT 01-MAR-2003 (TREMELrel. 23, Last sequence update)
DT 01-OCT-2003 (TREMELrel. 25, Last annotation update)
DE 13 days embryo head cDNA, RIKEN full-length enriched library,
DE clone:3110030A04 product:CED-6 protein homolog.
GN CSD6.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.

RESULT 10
09XG37

```

DT 01-JUN-2003 (TREMBLrel. 24, Last annotation update)
DE MHC class II beta 1 (Fragment)
GN SANA.
OS Salvelinus namaycush (lake trout).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei;
OC Protacanthopterygii; Salmoniformes; Salmonidae; Salvelinus.
OX NCBI_TaxID=8040;
RN [1]
RP SEQUENCE FROM N.A.
RA Dorschner M.O., Duris T., Phillips R.B.;
RT "Diversity of a Lake Trout Mhc Class II Gene.";
RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF130026; AAD20889.1; -.
DR GO; GO:0016021; C: integral to membrane; IEA.
DR GO; GO:0045012; F: MHC class II receptor activity; IEA.
DR GO; GO:0019884; P: antigen presentation, exogenous antigen; IEA.
DR GO; GO:0019886; P: antigen processing, exogenous antigen via M. . .; IEA.
DR GO; GO:0006955; P: immune response; IEA.
DR InterPro; IPR000353; MHC II beta.
DR Pfam; PF00969; MHC II beta_1.
DR ProDom; PD000328; MHC_II_beta_1.
KW Glycoprotein; MHC II; Transmembrane.
FT NON_TER 1
FT NON_TER 71
SQ SEQUENCE 71 AA; 8002 MW; E3095286582A9F2D CRC64;

Query Match 56.3%; Score 44.5; DB 7; Length 71;
Best Local Similarity 55.6%; Pred. No. 3.3; 2; Indels 3; Gaps 1;
Matches 10; Conservative 3; Mismatches 2; Indels 3; Gaps 1;

QY 2 QYIKANS---XFIGITEL 16
Db 14 EYIRFNSTVGKFGVGYTEL 31
::: ||| |||

RESULT 6
Q95IS2 PRELIMINARY; PRT; 85 AA.
ID Q95IS2;
AC Q95IS2;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-JUN-2003 (TREMBLrel. 24, Last annotation update)
DE MHC class II beta chain (Fragment).
OS Salmo salar (Atlantic salmon).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei;
OC Protacanthopterygii; Salmoniformes; Salmonidae; Salmo.
OX NCBI_TaxID=8030;
RN [1]
RP SEQUENCE FROM N.A.
RA Landry C., Bernatchez L.;
RT "Comparative analysis of population structure across environments and geographic scales at Major Histocompatibility Complex and microsatellite in Atlantic salmon (Salmo salar).";
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF273629; AAK1892.1; -.
DR GO; GO:0016021; C: integral to membrane; IEA.
DR GO; GO:0045012; F: MHC class II receptor activity; IEA.
DR GO; GO:0019884; P: antigen presentation, exogenous antigen; IEA.
DR GO; GO:0019886; P: antigen processing, exogenous antigen via M. . .; IEA.
DR GO; GO:0006955; P: immune response; IEA.
DR InterPro; IPR000353; MHC II beta.
DR Pfam; PF00969; MHC II beta_1.
DR ProDom; PD000328; MHC_II_beta_1.
KW MHC.
FT NON_TER 1
FT NON_TER 85
SQ SEQUENCE 85 AA; 9743 MW; 3214E01AD1B66AC5 CRC64;

Query Match 56.3%; Score 44.5; DB 7; Length 85;
Best Local Similarity 55.6%; Pred. No. 4; 3; Indels 2; Gaps 1;
Matches 10; Conservative 3; Mismatches 2; Indels 3; Gaps 1;

QY 2 QYIKANS---XFIGITEL 16
Db 14 EYIRFNSTVGKFGVGYTEL 31
::: ||| |||

RESULT 6
Q95IS2 PRELIMINARY; PRT; 85 AA.
ID Q95IS2;
AC Q95IS2;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-JUN-2003 (TREMBLrel. 24, Last annotation update)
DE MHC class II beta chain (Fragment).
OS Salmo salar (Atlantic salmon).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei;
OC Protacanthopterygii; Salmoniformes; Salmonidae; Salmo.
OX NCBI_TaxID=8030;
RN [1]
RP SEQUENCE FROM N.A.
RA Landry C., Bernatchez L.;
RT "Comparative analysis of population structure across environments and geographic scales at Major Histocompatibility Complex and microsatellite in Atlantic salmon (Salmo salar).";
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF273629; AAK1892.1; -.
DR GO; GO:0016021; C: integral to membrane; IEA.
DR GO; GO:0045012; F: MHC class II receptor activity; IEA.
DR GO; GO:0019884; P: antigen presentation, exogenous antigen; IEA.
DR GO; GO:0019886; P: antigen processing, exogenous antigen via M. . .; IEA.
DR GO; GO:0006955; P: immune response; IEA.
DR InterPro; IPR000353; MHC II beta.
DR Pfam; PF00969; MHC II beta_1.
DR ProDom; PD000328; MHC_II_beta_1.
KW MHC.
FT NON_TER 1
FT NON_TER 85
SQ SEQUENCE 85 AA; 9743 MW; 3214E01AD1B66AC5 CRC64;

Query Match 56.3%; Score 44.5; DB 7; Length 85;
Best Local Similarity 55.6%; Pred. No. 4; 3; Indels 2; Gaps 1;
Matches 10; Conservative 3; Mismatches 2; Indels 3; Gaps 1;

QY 2 QYIKANS---XFIGITEL 16
Db 33 EYIRFNSTVGKFGVGYTEL 50
::: ||| ||| |||

RESULT 8
Q95HX4 PRELIMINARY; PRT; 86 AA.
ID Q95HX4;
AC Q95HX4;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-JUN-2003 (TREMBLrel. 24, Last annotation update)
DE MHC class II B antigen (Fragment).
OS Salmo salar (Atlantic salmon).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei;
OC Protacanthopterygii; Salmoniformes; Salmonidae; Salmo.
OX NCBI_TaxID=8030;
RN [1]
RP SEQUENCE FROM N.A.
RA Langeforfs A., Lohm J., von Schantz T.;
RT "Allelic polymorphism in MHC class II B in four populations of Atlantic salmon (Salmo salar).";
RL Immunogenetics 53:329-336(2001).
DR EMBL; AF104377; AAL04009.1; -.
DR GO; GO:0016021; C: integral to membrane; IEA.

```

```

QY 2 QYIKANS---XFIGITEL 16
Db 33 EYIRFNSTVGKFGVGYTEL 50
::: ||| ||| |||

RESULT 7
Q95HY1 PRELIMINARY; PRT; 85 AA.
ID Q95HY1;
AC Q95HY1;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-JUN-2003 (TREMBLrel. 24, Last annotation update)
DE MHC class II B antigen (Fragment).
OS Salmo salar (Atlantic salmon).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei;
OC Protacanthopterygii; Salmoniformes; Salmonidae; Salmo.
OX NCBI_TaxID=8030;
RN [1]
RP SEQUENCE FROM N.A.
RA Langeforfs A., Lohm J., von Schantz T.;
RT "Allelic polymorphism in MHC class II B in four populations of Atlantic salmon (Salmo salar).";
RL Immunogenetics 53:329-336(2001).
DR EMBL; AF104370; AAL04002.1; -.
DR GO; GO:0016021; C: integral to membrane; IEA.
DR GO; GO:0045012; F: MHC class II receptor activity; IEA.
DR GO; GO:0019884; P: antigen presentation, exogenous antigen; IEA.
DR GO; GO:0019886; P: antigen processing, exogenous antigen via M. . .; IEA.
DR GO; GO:0006955; P: immune response; IEA.
DR InterPro; IPR000353; MHC II beta.
DR Pfam; PF00969; MHC II beta_1.
DR ProDom; PD000328; MHC_II_beta_1.
KW MHC.
FT NON_TER 1
FT NON_TER 85
SQ SEQUENCE 85 AA; 9723 MW; 27C9F7931F1F01C4 CRC64;

Query Match 56.3%; Score 44.5; DB 7; Length 85;
Best Local Similarity 55.6%; Pred. No. 4; 3; Indels 2; Gaps 1;
Matches 10; Conservative 3; Mismatches 2; Indels 3; Gaps 1;

QY 2 QYIKANS---XFIGITEL 16
Db 33 EYIRFNSTVGKFGVGYTEL 50
::: ||| ||| |||

RESULT 8
Q95HX4 PRELIMINARY; PRT; 86 AA.
ID Q95HX4;
AC Q95HX4;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-JUN-2003 (TREMBLrel. 24, Last annotation update)
DE MHC class II B antigen (Fragment).
OS Salmo salar (Atlantic salmon).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei;
OC Protacanthopterygii; Salmoniformes; Salmonidae; Salmo.
OX NCBI_TaxID=8030;
RN [1]
RP SEQUENCE FROM N.A.
RA Langeforfs A., Lohm J., von Schantz T.;
RT "Allelic polymorphism in MHC class II B in four populations of Atlantic salmon (Salmo salar).";
RL Immunogenetics 53:329-336(2001).
DR EMBL; AF104377; AAL04009.1; -.
DR GO; GO:0016021; C: integral to membrane; IEA.

```

Best Local Similarity 100.0%; Pred. NO. 4.3e-05;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MOVKANSKFIGITEL 16
Db 830 MQYKANSKFIGITEL 845

RESULT 2
Q7VQH3 PRELIMINARY; PRT; 447 AA.
AC Q7VQH3;
DT 01-OCT-2003 (TREMBlrel. 25, Created)
DT 01-OCT-2003 (TREMBlrel. 25, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE Enolase (EC 4.2.1.11).
GN ENO OR BFL157.
OS Candidatus Blochmannia floridanus.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; ant endosymbionts; Candidatus Blochmannia.
OX NCBI_TaxID=203907;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=22784745; PubMed=12886019;
RA Gil R., Silva F.J., Zientz E., Delmotte F., Gonzalez-Candelas F.,
RA Latorre A., Rausell C., Kamerbeek J., Gadau J., Hoelldobler B.,
RA van Ham R.C.H.J., Gross R., Moya A.;
RT "The genome sequence of Blochmannia floridanus: comparative analysis
of reduced genomes";
RT Proc. Natl. Acad. Sci. U.S.A. 100:9388-9393(2003).
RL EMBL; BX248584; CAD83678.1; -.
KW Lysase; Complete proteome.
SQ SEQUENCE 447 AA; 49005 MW; 465B69C32737CA4 CRC64;

Query Match 58.2%; Score 46; DB 16; Length 447;
Best Local Similarity 46.7%; Pred. NO. 11; Mismatches 5; Indels 0; Gaps 0;
Matches 7; Conservative 5;

QY 2 QYIKANSKFIGITEL 16
Db 416 EFLKDNSKFIGVNI 430

RESULT 3
Q835J8 PRELIMINARY; PRT; 880 AA.
AC Q835J8;
DT 01-JUN-2003 (TREMBlrel. 24, Created)
DT 01-JUN-2003 (TREMBlrel. 24, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE Alanine-tRNA synthetase.
GN ALAS OR EFL379.
OS Enterococcus faecalis (Streptococcus faecalis).
OC Bacteria; Firmicutes; Lactobacilliales; Enterococcaceae; Enterococcus.
OX NCBI_TaxID=1351;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=V563 / ATCC 700802;
RX MEDLINE=22550857; PubMed=12663927;
RA Paulsen I.T., Banerjee L., Myers G.S.A., Nelson K.E., Seshadri R.,
RA Read T.D., Fouts D.E., Eisen J.A., Gill S.R., Heidelberg J.F.,
RA Tettelin H., Dodson R.J., Umayam L., Brinkac L., Beanan M.,
RA Daugherty S., DeBoy R.T., Durkin S., Kolonay J., Madupu R., Nelson W.,
RA Vamathevan J., Tran B., Upton J., Hansen T., Shetty J., Khouri H.,
RA Utterback T., Radune D., Ketchum K.A., Dougherty B.A., Fraser C.M.;
RT "Role of mobile DNA in the evolution of vancomycin-resistant
Enterococcus faecalis";
RL Science 299:2071-2074(2003).
RL EMBL; AE016951; AAO81170.1; -.
DR TIGR; EF1379; -.
DR GO; GO:0004813; Alanine-tRNA ligase activity; IEA.
DR GO; GO:0005524; F-ATP binding; IEA.
DR GO; GO:0004197; F-cysteine-type endopeptidase activity; IEA.

DR GO; GO:0003676; F-nucleic acid binding; IEA.
DR GO; GO:0006419; Palanyl-CRNA aminoacylation; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR003156; DHHA1.
DR InterPro; IPR000169; SHprot_acsite.
DR InterPro; IPR002318; tRNA-synt_2c.
DR InterPro; IPR006193; tRNA_synt_Ala.
DR Pfam; PF02272; DHHA1; 1.
DR Pfam; PF01411; tRNA-synt_2c; 1.
DR PRINTS; PRO0980; TRNASINTHALA.
DR TIGRFAMs; TIGR00344; alas; 1.
DR PROSITE; PS0860; AA TRNA LIGASE II ALA; 1.
DR PROSITE; PS0863; THIOL PROTEASE HIS; 1.
KW Aminoacyl-tRNA synthetase; Complete proteome.
SQ SEQUENCE 880 AA; 97887 MW; EF23BD1229B6683F CRC64;

Query Match 58.2%; Score 46; DB 16; Length 880;
Best Local Similarity 69.2%; Pred. NO. 22; Mismatches 1; Indels 0; Gaps 0;
Matches 9; Conservative 1;

QY 4 IKANSKFIGITEL 16
Db 455 IKVESKFIGITEL 467

RESULT 4
Q31585 PRELIMINARY; PRT; 60 AA.
AC Q31585;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-JUN-2003 (TREMBlrel. 24, Last annotation update)
DE (DB02) MHC class II beta.1 (fragment).
OS Salmo salar (Atlantic salmon).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei;
OC Protacanthopterygii; Salmoniformes; Salmonidae; Salmo.
OX NCBI_TaxID=8030;
RN [1]
RP SEQUENCE FROM N.A.
RA Grimholt U., Olsaker I., de Vries Lindstrom C., Lie O.;
RT "A study of polymorphism in the MHC class II beta.1 and MHC class I
alpha.2 domain exons of Atlantic salmon (Salmo salar).";
RL Submitted (OCT-1993) to the EMBL/GenBank/DBJ databases.
DR EMBL; L24953; AAA49597.1; -.
DR GO; GO:0015021; C:integral to membrane; IEA.
DR GO; GO:0045012; P:MHC class II receptor activity; IEA.
DR GO; GO:0019884; P:antigen presentation, exogenous antigen; IEA.
DR GO; GO:0019886; P:antigen processing, exogenous antigen via M...; IEA.
DR GO; GO:0006955; P:immune response; IEA.
DR InterPro; IPR000353; MHC II beta.
DR Pfam; PF00969; MHC II beta; 1.
DR ProDom; PD000328; MHC_II_beta; 1.
KW Glycoprotein; MHC II; Transmembrane.
FT NON_TER 1
FT NON_TER 60
SQ SEQUENCE 60 AA; 6776 MW; 5D4F3449060940E2 CRC64;

Query Match 56.3%; Score 44.5; DB 7; Length 60;
Best Local Similarity 55.6%; Pred. NO. 2.8;
Matches 10; Conservative 3; Mismatches 2; Indels 3; Gaps 1;

QY 2 QYIKANSKFIGITEL 16
Db 16 EXIRFNSKFIGITEL 33

RESULT 5
Q9XRJ9 PRELIMINARY; PRT; 71 AA.
ID Q9XRJ9
AC Q9XRJ9;
DT 01-NOV-1999 (TREMBlrel. 12, Created)
DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)

Result No.	Query %			DB	ID	Description
	Score	Match	Length			
1	79	100.0	1310	2	Q93N27	clostridium
2	46	58.2	447	16	Q7VOH3	Q93WH3 candidatus
3	46	58.2	880	16	Q835J8	enterococcus
4	44.5	56.3	60	7	Q1S8S5	salmo salar
5	44.5	56.3	71	7	Q9XRJ9	salvelinus
6	44.5	56.3	85	7	Q95IS2	salmo salar
7	44.5	56.3	85	7	Q95HY1	salmo salar
8	44.5	56.3	86	7	Q95HX4	salmo salar
9	44.5	56.3	244	7	Q1S590	salmo salar
10	44	55.7	546	10	Q9XG37	guillardia
11	43.5	55.1	67	7	Q1S578	salmo salar
12	43	54.4	180	16	Q9CF66	lactococcus
13	43	54.4	250	9	Q9NCL7	streptococcus
14	43	54.4	252	9	Q9XJEB	lactococcus
15	43	54.4	291	11	Q9CRV4	mus musculus
16	43	54.4	304	11	Q8K2A1	mus musculus

```
; NAME/KEY: PEPTIDE
; LOCATION: (1)..(10)
; OTHER INFORMATION: Amino acid sequence 1-10 of the human GnRH hormone
; NAME/KEY: PEPTIDE
; LOCATION: (11)..(16)
; OTHER INFORMATION: Spacer peptide
; NAME/KEY: PEPTIDE
; LOCATION: (17)..(31)
; OTHER INFORMATION: Amino acid sequence 830-844 of the Tetanus toxoid precursor
; NAME/KEY: PEPTIDE
; LOCATION: (32)..(37)
; OTHER INFORMATION: Spacer peptide
; NAME/KEY: PEPTIDE
; LOCATION: (38)..(46)
; OTHER INFORMATION: Amino acid sequence 2-10 of the human GnRH hormone
US-09-848-834A-19

Query Match          96.2%; Score 76; DB 9; Length 46;
Best Local Similarity 93.8%; Pred. No. 1.1e-06;
Matches 15; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 MOYIKANSKFIGITEL 16
Db 16 LQYIKANSKFIGITEL 31

RESULT 13
US-10-297-942-8
; Sequence 8, Application US/10297942
; Publication No. US20030185816A1
; GENERAL INFORMATION:
; APPLICANT: Ferring BV
; TITLE OF INVENTION: Solubilised Protein Vaccines
; FILE REFERENCE: P88445USO
; CURRENT APPLICATION NUMBER: US/10/297,942
; CURRENT FILING DATE: 2003-04-21
; PRIOR APPLICATION NUMBER: PCT/DK01/00431
; PRIOR FILING DATE: 2001-10-16
; PRIOR APPLICATION NUMBER: DK PA 2000 00966
; PRIOR FILING DATE: 2000-06-21
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 8
; LENGTH: 158
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-297-942-8

Query Match          96.2%; Score 76; DB 14; Length 158;
Best Local Similarity 93.8%; Pred. No. 4.5e-06;
Matches 15; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 MOYIKANSKFIGITEL 16
Db 76 LQYIKANSKFIGITEL 91

RESULT 14
US-09-943-548-9
; Sequence 9, Application US/09943548
; Patent No. US20020042364A1
; GENERAL INFORMATION:
; APPLICANT: Rittershaus, Charles W.
; APPLICANT: Thomas, Lawrence J.
; TITLE OF INVENTION: MODULATION OF CHOLESTERYL ESTER TRANSFER PROTEIN (CETP) ACTIVITY
; FILE REFERENCE: TCS-411.1P US-1; TCS-411.1P US-2
; CURRENT APPLICATION NUMBER: US/09/943,548
; CURRENT FILING DATE: 2001-08-30
; PRIOR APPLICATION NUMBER: 08/432,483
; PRIOR FILING DATE: 1995-05-01
; PRIOR APPLICATION NUMBER: PCT/US96/06147
```

```
; PRIOR APPLICATION NUMBER: 08/945,289
; PRIOR FILING DATE: 1997-10-17
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 9
; LENGTH: 50
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: vaccine peptide of the invention
US-09-943-548-9

Query Match          94.9%; Score 75; DB 9; Length 50;
Best Local Similarity 100.0%; Pred. No. 1.9e-06;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MOYIKANSKFIGITE 15
Db 1 MOYIKANSKFIGITE 15

RESULT 15
US-10-339-522-9
; Sequence 9, Application US/10339522
; Publication No. US20030108559A1
; GENERAL INFORMATION:
; APPLICANT: Rittershaus, Charles W.
; APPLICANT: Thomas, Lawrence J.
; TITLE OF INVENTION: MODULATION OF CHOLESTERYL ESTER TRANSFER PROTEIN (CETP) ACTIVITY
; FILE REFERENCE: TCS-411.1P US-3
; CURRENT APPLICATION NUMBER: US/10/339,522
; CURRENT FILING DATE: 2003-01-08
; PRIOR APPLICATION NUMBER: 08/432,483
; PRIOR FILING DATE: 1995-05-01
; PRIOR APPLICATION NUMBER: PCT/US96/06147
; PRIOR FILING DATE: 1996-05-01
; PRIOR APPLICATION NUMBER: 08/945,289
; PRIOR FILING DATE: 1997-10-17
; PRIOR APPLICATION NUMBER: 09/943,334
; PRIOR FILING DATE: 2001-08-30
; PRIOR APPLICATION NUMBER: 09/943,548
; PRIOR FILING DATE: 2001-08-30
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 9
; LENGTH: 50
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: vaccine peptide of the invention
US-10-339-522-9

Query Match          94.9%; Score 75; DB 14; Length 50;
Best Local Similarity 100.0%; Pred. No. 1.9e-06;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MOYIKANSKFIGITE 15
Db 1 MOYIKANSKFIGITE 15

Search completed: March 10, 2004, 10:25:47
Job time : 12.5759 secs
```


RESULT 9

US-10-241-596-141
; Sequence 141, Application US/10241596
; Publication No. US20030166238A1
; GENERAL INFORMATION:
; APPLICANT: Microbiological Research Authority
; APPLICANT: The Speywood Laboratory Limited
; TITLE OF INVENTION: Recombinant Toxin Fragments
; FILE REFERENCE: 1581.0130003
; CURRENT APPLICATION NUMBER: US/10/241,596
; CURRENT FILING DATE: 2002-09-12
; PRIOR APPLICATION NUMBER: US 09/255,829
; PRIOR FILING DATE: 1999-02-23
; PRIOR APPLICATION NUMBER: US 09/242,689
; PRIOR FILING DATE: 1999-02-23
; PRIOR APPLICATION NUMBER: PCT/GB97/02273
; PRIOR FILING DATE: 1997-08-22
; PRIOR APPLICATION NUMBER: US 08/782,893
; PRIOR FILING DATE: 1996-12-27
; PRIOR APPLICATION NUMBER: GB 9625996.5
; PRIOR FILING DATE: 1996-12-13
; PRIOR APPLICATION NUMBER: GB 9617671.4
; PRIOR FILING DATE: 1996-08-23
; NUMBER OF SEQ ID NOS: 175
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 141
; LENGTH: 1315
; TYPE: PRT
; ORGANISM: Clostridium tetani
US-10-241-596-141

Query Match 100.0%; Score 79; DB 14; Length 1315;
Best Local Similarity 100.0%; Pred. No. 1.4e-05;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MQYIKANSKFIGITEL 16
| | | | | | | | | | | | | | | | | | | | | |
Db 829 MQYIKANSKFIGITEL 844

RESULT 10

US-10-452-024-145
; Sequence 145, Application US/10452024
; Publication No. US20040013687A1
; GENERAL INFORMATION:
; APPLICANT: Simpson, Lance
; APPLICANT: Park, Jung-Beak
; APPLICANT: Maksymowich, Andrew
; TITLE OF INVENTION: Compositions and Methods For Transsepithelial Molecular Transport
; FILE REFERENCE: 9855-9601
; CURRENT APPLICATION NUMBER: US/10/452,024
; CURRENT FILING DATE: 2003-06-02
; PRIOR APPLICATION NUMBER: 60/384,949
; PRIOR FILING DATE: 2002-05-31
; NUMBER OF SEQ ID NOS: 188
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 145
; LENGTH: 1315
; TYPE: PRT
; ORGANISM: Clostridium tetani
US-10-452-024-145

Query Match 100.0%; Score 79; DB 15; Length 1315;
Best Local Similarity 100.0%; Pred. No. 1.4e-05;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MQYIKANSKFIGITEL 16
| | | | | | | | | | | | | | | | | | | | | |
Db 829 MQYIKANSKFIGITEL 844

RESULT 11

US-09-848-834A-15

; Sequence 15, Application US/09848834A
; Patent No. US20020076416A1
; GENERAL INFORMATION:
; APPLICANT: Aphton Corporation
; TITLE OF INVENTION: Chimeric Peptide Immunogens
; FILE REFERENCE: 1102865-0047
; CURRENT APPLICATION NUMBER: US/09/848,834A
; CURRENT FILING DATE: 2001-05-04
; PRIOR APPLICATION NUMBER: 60/202,328
; PRIOR FILING DATE: 2000-05-05
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 15
; LENGTH: 31
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Chimeric peptide consisting of amino acid sequence 1-10 of the (C
; OTHER INFORMATION: RH hormone linked by a spacer to amino sequence 830-844 of the (1
; OTHER INFORMATION: tanus toxoid precursor (Tentoxylisin)
; NAME/KEY: MOD_RES
; LOCATION: (1)..(1)
; OTHER INFORMATION: Pyroglutamic acid or 5-oxoproline
; NAME/KEY: PEPTIDE
; LOCATION: (1)..(10)
; OTHER INFORMATION: Amino acid sequence 1-10 of the human GnRH hormone
; NAME/KEY: PEPTIDE
; LOCATION: (11)..(16)
; OTHER INFORMATION: Spacer peptide
; NAME/KEY: PEPTIDE
; LOCATION: (17)..(31)
; OTHER INFORMATION: Amino acid sequence 830-844 of the Tetanus toxoid precursor
; OTHER INFORMATION: (Tentoxylisin)
US-09-848-834A-15

Query Match 96.2%; Score 76; DB 9; Length 31;
Best Local Similarity 93.8%; Pred. No. 7.1e-07;
Matches 15; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MQYIKANSKFIGITEL 16
| | | | | | | | | | | | | | | | | | | | | |
Db 16 MQYIKANSKFIGITEL 31

RESULT 12

US-09-848-834A-19
; Sequence 19, Application US/09848834A
; Patent No. US20020076416A1
; GENERAL INFORMATION:
; APPLICANT: Aphton Corporation
; TITLE OF INVENTION: Chimeric Peptide Immunogens
; FILE REFERENCE: 1102865-0047
; CURRENT APPLICATION NUMBER: US/09/848,834A
; CURRENT FILING DATE: 2001-05-04
; PRIOR APPLICATION NUMBER: 60/202,328
; PRIOR FILING DATE: 2000-05-05
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 19
; LENGTH: 46
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Chimeric peptide consisting of amino acid sequence 1-10 of human
; OTHER INFORMATION: GnRH linked by a spacer to amino acid sequence 830-844 of Tetan
; OTHER INFORMATION: toxoid precursor (Tentoxylisin) linked by a spacer to amino aci
; OTHER INFORMATION: sequence 1-10 of GnRH
; NAME/KEY: MOD_RES
; LOCATION: (1)..(1)
; OTHER INFORMATION: Pyroglutamic acid or 5-oxoproline
; NAME/KEY: MOD_RES
; LOCATION: (46)..(46)
; OTHER INFORMATION: Amidated glycine or glycynamide

APPLICANT: Microbiological Research Authority
APPLICANT: The Speywood Laboratory Limited
TITLE OF INVENTION: Recombinant Toxin Fragments
FILE REFERENCE: 1581.0130003
CURRENT FILING DATE: 2002-09-12
CURRENT APPLICATION NUMBER: US 09/355,829
PRIOR FILING DATE: 1999-02-23
PRIOR APPLICATION NUMBER: US 09/242,689
PRIOR FILING DATE: 1999-02-23
PRIOR APPLICATION NUMBER: PCT/GB97/02273
PRIOR FILING DATE: 1997-08-22
PRIOR APPLICATION NUMBER: US 08/782,893
PRIOR FILING DATE: 1996-12-27
PRIOR APPLICATION NUMBER: GB 9625996.5
PRIOR FILING DATE: 1996-12-13
PRIOR APPLICATION NUMBER: GB 9617671.4
PRIOR FILING DATE: 1996-08-23
NUMBER OF SEQ ID NOS: 175
SOFTWARE: Patent in version 3.1
SEQ ID NO 145
LENGTH: 872
TYPE: PRT
ORGANISM: Clostridium tetani
US-10-241-596-145

Query Match 100.0%; Score 79; DB 14; Length 872;
Best Local Similarity 100.0%; Pred. No. 8.7e-06;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MOYIKANSKFIGITEL 16
DB 822 MOYIKANSKFIGITEL 837

RESULT 6
US-10-241-596-143
Sequence 143, Application US/10241596
Publication No. US20030166238A1
GENERAL INFORMATION:
APPLICANT: Microbiological Research Authority
APPLICANT: The Speywood Laboratory Limited
TITLE OF INVENTION: Recombinant Toxin Fragments
FILE REFERENCE: 1581.0130003
CURRENT FILING DATE: 2002-09-12
CURRENT APPLICATION NUMBER: US/10/241,596
PRIOR FILING DATE: 1999-02-23
PRIOR APPLICATION NUMBER: US 09/255,829
PRIOR FILING DATE: 1999-02-23
PRIOR APPLICATION NUMBER: US 09/242,689
PRIOR FILING DATE: 1999-02-23
PRIOR APPLICATION NUMBER: PCT/GB97/02273
PRIOR FILING DATE: 1997-08-22
PRIOR APPLICATION NUMBER: US 08/782,893
PRIOR FILING DATE: 1996-12-27
PRIOR APPLICATION NUMBER: GB 9625996.5
PRIOR FILING DATE: 1996-12-13
PRIOR APPLICATION NUMBER: GB 9617671.4
PRIOR FILING DATE: 1996-08-23
NUMBER OF SEQ ID NOS: 175
SOFTWARE: Patent in version 3.1
SEQ ID NO 143
LENGTH: 879
TYPE: PRT
ORGANISM: Clostridium tetani
US-10-241-596-143

Query Match 100.0%; Score 79; DB 14; Length 879;
Best Local Similarity 100.0%; Pred. No. 8.7e-06;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MOYIKANSKFIGITEL 16
DB 829 MOYIKANSKFIGITEL 844

RESULT 7
US-10-241-596-147
Sequence 147, Application US/10241596
Publication No. US20030166238A1
GENERAL INFORMATION:
APPLICANT: Microbiological Research Authority
APPLICANT: The Speywood Laboratory Limited
TITLE OF INVENTION: Recombinant Toxin Fragments
FILE REFERENCE: 1581.0130003
CURRENT FILING DATE: 2002-09-12
CURRENT APPLICATION NUMBER: US 09/255,829
PRIOR FILING DATE: 1999-02-23
PRIOR APPLICATION NUMBER: US 09/242,689
PRIOR FILING DATE: 1999-02-23
PRIOR APPLICATION NUMBER: PCT/GB97/02273
PRIOR FILING DATE: 1997-08-22
PRIOR APPLICATION NUMBER: US 08/782,893
PRIOR FILING DATE: 1996-12-27
PRIOR APPLICATION NUMBER: GB 9625996.5
PRIOR FILING DATE: 1996-12-13
PRIOR APPLICATION NUMBER: GB 9617671.4
PRIOR FILING DATE: 1996-08-23
NUMBER OF SEQ ID NOS: 175
SOFTWARE: Patent in version 3.1
SEQ ID NO 147
LENGTH: 887
TYPE: PRT
ORGANISM: Clostridium tetani
US-10-241-596-147

Query Match 100.0%; Score 79; DB 14; Length 887;
Best Local Similarity 100.0%; Pred. No. 8.8e-06;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MOYIKANSKFIGITEL 16
DB 829 MOYIKANSKFIGITEL 844

RESULT 8
US-10-452-024-149
Sequence 149, Application US/10452024
Publication No. US20040013687A1
GENERAL INFORMATION:
APPLICANT: Simpson, Lance
APPLICANT: Park, Jung-Beak
APPLICANT: Maksymowich, Andrew
TITLE OF INVENTION: Compositions and Methods For Transsepithelial Molecular Transport
FILE REFERENCE: 9855-96U1
CURRENT APPLICATION NUMBER: US/10/452,024
CURRENT FILING DATE: 2003-06-02
PRIOR APPLICATION NUMBER: 60/384,949
PRIOR FILING DATE: 2002-05-31
NUMBER OF SEQ ID NOS: 188
SOFTWARE: Patent in version 3.2
SEQ ID NO 149
LENGTH: 1310
TYPE: PRT
ORGANISM: clostridium tetani
US-10-452-024-149

Query Match 100.0%; Score 79; DB 15; Length 1310;
Best Local Similarity 100.0%; Pred. No. 1.4e-05;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MOYIKANSKFIGITEL 16
DB 830 MOYIKANSKFIGITEL 845

```

;
; GENERAL INFORMATION:
; APPLICANT: Hickey, William F.
; TITLE OF INVENTION: Proinsulin Peptide Compounds for Detecting and
; Treating Type I Diabetes
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD
; STREET: 60 State Street, suite 510
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109-1875
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/346,563
; FILING DATE: 16-Jan-2003
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/472,704
; FILING DATE: 06-Jun-1995
; APPLICATION NUMBER: US 08/272,220
; FILING DATE: 08-JULY-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: DeConti, Giulio A., Jr.
; REGISTRATION NUMBER: 31,503
; REFERENCE/DOCKET NUMBER: DCI-092
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)227-7400
; TELEFAX: (617)227-5941
; INFORMATION FOR SEQ ID NO: 23:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 17 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; FRAGMENT TYPE: internal
; SEQUENCE DESCRIPTION: SEQ ID NO: 23:
US-10-346-563-23

Query Match 100.0%; Score 79; DB 15; Length 17;
Best Local Similarity 100.0%; Pred. No. 1e-07;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MOYIKANSKFIGITEL 16
Db 2 MOYIKANSKFIGITEL 17

RESULT 3
US-10-321-717-23
; Sequence 23, Application US/10321717
; Publication No. US20040002113A1
; GENERAL INFORMATION:
; APPLICANT: Hickey, William F.
; TITLE OF INVENTION: Detection and Treatment Methods for
; Treating Type I Diabetes
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD
; STREET: 60 State Street, suite 510
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109-1875
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible

```

```

;
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/321,717
; FILING DATE: 17-Dec-2002
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/472,701
; FILING DATE: <Unknown>
; APPLICATION NUMBER: US 08/272,220
; FILING DATE: 08-JULY-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: DeConti, Giulio A., Jr.
; REGISTRATION NUMBER: 31,503
; REFERENCE/DOCKET NUMBER: DCI-092DV
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)227-7400
; TELEFAX: (617)227-5941
; INFORMATION FOR SEQ ID NO: 23:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 17 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; FRAGMENT TYPE: internal
; SEQUENCE DESCRIPTION: SEQ ID NO: 23:
US-10-321-717-23

Query Match 100.0%; Score 79; DB 15; Length 17;
Best Local Similarity 100.0%; Pred. No. 1e-07;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MOYIKANSKFIGITEL 16
Db 2 MOYIKANSKFIGITEL 17

RESULT 4
US-10-452-024-177
; Sequence 177, Application US/10452024
; Publication No. US20040013687A1
; GENERAL INFORMATION:
; APPLICANT: Simpson, Lance
; APPLICANT: Park, Jung-Beak
; APPLICANT: Maksymowych, Andrew
; TITLE OF INVENTION: Compositions and Methods For Transsepithelial Molecular Transport
; FILE REFERENCE: 9855-96U1
; CURRENT APPLICATION NUMBER: US/10/452,024
; CURRENT FILING DATE: 2003-06-02
; PRIOR APPLICATION NUMBER: 60/384,949
; PRIOR FILING DATE: 2002-05-31
; NUMBER OF SEQ ID NOS: 188
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 177
; LENGTH: 573
; TYPE: PRT
; ORGANISM: Clostridium tetani
US-10-452-024-177

Query Match 100.0%; Score 79; DB 15; Length 573;
Best Local Similarity 100.0%; Pred. No. 5.4e-06;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MOYIKANSKFIGITEL 16
Db 87 MOYIKANSKFIGITEL 102

RESULT 5
US-10-241-596-145
; Sequence 145, Application US/10241596
; Publication No. US20030166238A1
; GENERAL INFORMATION:

```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 10, 2004, 09:16:59 ; Search time 12.5759 Seconds
(without alignments)
268.645 Million cell updates/sec

Title: US-09-848-834A-2

Perfect score: 79
Sequence: 1 MQYIKANSKFIGITEL 16

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 809742 seqs, 211153259 residues

Total number of hits satisfying chosen parameters: 809742

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications AA:
1: /cgn2_6/prodata/2/pubpaa/US07_PUBCOMB.pep:
2: /cgn2_6/prodata/2/pubpaa/PCT_NEW_PUB.pep:
3: /cgn2_6/prodata/2/pubpaa/US06_NEW_PUB.pep:
4: /cgn2_6/prodata/2/pubpaa/US06_PUBCOMB.pep:
5: /cgn2_6/prodata/2/pubpaa/US07_NEW_PUB.pep:
6: /cgn2_6/prodata/2/pubpaa/PCTUS_PUBCOMB.pep:
7: /cgn2_6/prodata/2/pubpaa/US08_NEW_PUB.pep:
8: /cgn2_6/prodata/2/pubpaa/US08_PUBCOMB.pep:
9: /cgn2_6/prodata/2/pubpaa/US09A_PUBCOMB.pep:
10: /cgn2_6/prodata/2/pubpaa/US09B_PUBCOMB.pep:
11: /cgn2_6/prodata/2/pubpaa/US09C_PUBCOMB.pep:
12: /cgn2_6/prodata/2/pubpaa/US09_NEW_PUB.pep:
13: /cgn2_6/prodata/2/pubpaa/US10A_PUBCOMB.pep:
14: /cgn2_6/prodata/2/pubpaa/US10B_PUBCOMB.pep:
15: /cgn2_6/prodata/2/pubpaa/US10C_PUBCOMB.pep:
16: /cgn2_6/prodata/2/pubpaa/US10_NEW_PUB.pep:
17: /cgn2_6/prodata/2/pubpaa/US60_NEW_PUB.pep:
18: /cgn2_6/prodata/2/pubpaa/US60_PUBCOMB.pep:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	79	100.0	16	9	US-09-848-834A-2
2	79	100.0	17	15	US-10-346-563-23
3	79	100.0	17	15	US-10-321-717-23
4	79	100.0	573	15	US-10-452-024-177
5	79	100.0	872	14	US-10-241-596-145
6	79	100.0	879	14	US-10-241-596-143
7	79	100.0	887	14	US-10-241-596-147
8	79	100.0	1310	15	US-10-452-024-149
9	79	100.0	1315	14	US-10-241-596-141
10	79	100.0	1315	15	US-10-452-024-145
11	76	96.2	31	9	US-09-848-834A-15
12	76	96.2	46	9	US-09-848-834A-19
13	76	96.2	158	14	US-10-297-942-8
14	75	94.9	50	9	US-09-943-548-9
15	75	94.9	50	14	US-10-339-522-9

16	74	93.7	15	9	US-09-862-849-2	Sequence 2, Appli
17	74	93.7	15	9	US-09-785-215-4	Sequence 4, Appli
18	74	93.7	15	10	US-09-405-986-1	Sequence 1, Appli
19	74	93.7	15	14	US-10-204-362-4	Sequence 4, Appli
20	74	93.7	15	14	US-10-223-711-7	Sequence 7, Appli
21	74	93.7	15	14	US-10-237-656-13	Sequence 13, Appli
22	74	93.7	15	14	US-10-223-809A-4	Sequence 4, Appli
23	74	93.7	15	14	US-10-261-446-19	Sequence 19, Appli
24	74	93.7	15	14	US-10-239-313A-618	Sequence 618, App
25	74	93.7	15	14	US-10-295-074-3	Sequence 3, Appli
26	74	93.7	15	15	US-10-372-111-7	Sequence 7, Appli
27	74	93.7	16	10	US-09-930-915A-64	Sequence 64, Appli
28	74	93.7	16	14	US-10-082-01A-145	Sequence 145, App
29	74	93.7	16	14	US-10-372-076-175	Sequence 175, App
30	74	93.7	17	10	US-09-865-294-3	Sequence 3, Appli
31	74	93.7	17	14	US-10-239-313A-619	Sequence 619, App
32	74	93.7	17	15	US-10-411-544-5	Sequence 5, Appli
33	74	93.7	19	14	US-10-239-313A-620	Sequence 620, App
34	74	93.7	27	14	US-10-076-674-7	Sequence 7, Appli
35	74	93.7	27	15	US-10-355-161A-7	Sequence 7, Appli
36	74	93.7	28	9	US-09-848-834A-11	Sequence 11, Appli
37	74	93.7	31	14	US-10-237-656-1	Sequence 1, Appli
38	74	93.7	32	14	US-10-237-656-5	Sequence 5, Appli
39	74	93.7	50	9	US-09-943-548-8	Sequence 8, Appli
40	74	93.7	50	14	US-10-339-522-8	Sequence 8, Appli
41	74	93.7	137	14	US-10-237-656-3	Sequence 3, Appli
42	74	93.7	158	14	US-10-297-942-4	Sequence 4, Appli
43	74	93.7	158	14	US-10-297-942-6	Sequence 6, Appli
44	74	93.7	158	14	US-10-297-942-14	Sequence 14, Appli
45	74	93.7	158	14	US-10-297-942-16	Sequence 16, Appli

ALIGNMENTS

RESULT 1

US-09-848-834A-2
; Sequence 2, Application US/09848834A
; Patent No. US20020076416A1
; GENERAL INFORMATION:
; APPLICANT: Aphton Corporation
; TITLE OF INVENTION: Chimeric Peptide Immunogens
; FILE REFERENCE: 1102865-0047
; CURRENT APPLICATION NUMBER: US/09/848,834A
; CURRENT FILING DATE: 2001-05-04
; PRIOR APPLICATION NUMBER: 60/202,328
; PRIOR FILING DATE: 2000-05-05
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: Patent version 3.0
; SEQ ID NO 2
; LENGTH: 16
; TYPE: PRT
; ORGANISM: Tetanus bacillus
; FEATURE:
; NAME/KEY: PEPTIDE
; LOCATION: (1)..(16)
; OTHER INFORMATION: Amino acid sequence 829-844 of the Tetanus
; OTHER INFORMATION: Toxoid Precursor (Tetoxylisin)
US-09-848-834A-2

Query Match 100.0%; Score 79; DB 9; Length 16;
Best Local Similarity 100.0%; Pred. No. 9.5e-08;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MQYIKANSKFIGITEL 16

Db 1 MQYIKANSKFIGITEL 16

RESULT 2

US-10-346-563-23
; Sequence 23, Application US/10346563
; Publication No. US20030220229A1


```

DR GO: GO:0003793; F:defense/immunity protein activity; IEA.
DR InterPro: IPR003067; Crmsprzoite.
DR InterPro: IPR000884; TSPI.
DR Pfam: PF00090; tsp.1; 1.
DR PRINTS: PR01303; CRMSPRZOITE.
DR SMART: SM00209; TSPI; 1.
DR PROSITE: PS50092; TSPI; 1.
DR NON_TER 1
FT NON_TER 1
FT NON_TER 80
SQ SEQUENCE 80 AA; 9043 MW; 17282B319AB508B3 CRC64;

Query Match      85.3%; Score 81; DB 5; Length 80;
Best Local Similarity 89.5%; Pred. No. 3.5e-05;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 EKXIAMEKASSVFNVNS 20
DB 62 EKXICKMEKCSSVFNVNS 80

RESULT 13
Q9U0Q4 PRELIMINARY; PRT; 80 AA.
ID Q9U0Q4
AC Q9U0Q4;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Circumsporozoite protein (Fragment).
GN CS.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RC STRAIN=B1;
RA de Stricker K., Vuust J., Jepsen S., Oeuvray C., Theisen M.;
RT "Sequence variation in the non-repeat region of the Plasmodium
RT falciparum glutamate rich protein (GLURP) from Brazil, Senegalese, and
RT Burmese field isolates and from laboratory strains.";
RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AJ269941; CAB64167.1; -.
DR GO: GO:0003793; F:defense/immunity protein activity; IEA.
DR InterPro: IPR003067; Crmsprzoite.
DR InterPro: IPR000884; TSPI.
DR Pfam: PF00090; tsp.1; 1.
DR PRINTS: PR01303; CRMSPRZOITE.
DR SMART: SM00209; TSPI; 1.
DR PROSITE: PS50092; TSPI; 1.
DR NON_TER 1
FT NON_TER 1
FT NON_TER 80
SQ SEQUENCE 80 AA; 9102 MW; A3283B70CEE50FDE CRC64;

Query Match      85.3%; Score 81; DB 5; Length 80;
Best Local Similarity 89.5%; Pred. No. 3.5e-05;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 EKXIAMEKASSVFNVNS 20
DB 62 EKXICKMEKCSSVFNVNS 80

RESULT 14
Q9TW00 PRELIMINARY; PRT; 80 AA.
ID Q9TW00
AC Q9TW00;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Circumsporozoite protein (fragment).
GN CS.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;

```

```

[1]
RN SEQUENCE FROM N.A.
RP STRAIN=FUP/SP, and MAD20;
RC de Stricker K., Vuust J., Jepsen S., Oeuvray C., Theisen M.;
RA "Sequence variation in the non-repeat region of the Plasmodium
RT falciparum glutamate rich protein (GLURP) from Brazil, Senegalese, and
RT Burmese field isolates and from laboratory strains.";
RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AJ269953; CAB64178.1; -.
DR GO: GO:0003793; F:defense/immunity protein activity; IEA.
DR InterPro: IPR003067; Crmsprzoite.
DR InterPro: IPR000884; TSPI.
DR Pfam: PF00090; tsp.1; 1.
DR PRINTS: PR01303; CRMSPRZOITE.
DR SMART: SM00209; TSPI; 1.
DR PROSITE: PS50092; TSPI; 1.
DR NON_TER 1
FT NON_TER 1
FT NON_TER 80
SQ SEQUENCE 80 AA; 9074 MW; 4614F9D18F1C0334 CRC64;

Query Match      85.3%; Score 81; DB 5; Length 80;
Best Local Similarity 89.5%; Pred. No. 3.5e-05;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 EKXIAMEKASSVFNVNS 20
DB 62 EKXICKMEKCSSVFNVNS 80

RESULT 15
Q9U0P6 PRELIMINARY; PRT; 80 AA.
ID Q9U0P6
AC Q9U0P6;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Circumsporozoite protein (fragment).
GN CS.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RC STRAIN=D4264;
RA de Stricker K., Vuust J., Jepsen S., Oeuvray C., Theisen M.;
RT "Sequence variation in the non-repeat region of the Plasmodium
RT falciparum glutamate rich protein (GLURP) from Brazil, Senegalese, and
RT Burmese field isolates and from laboratory strains.";
RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AJ269963; CAB64182.1; -.
DR GO: GO:0003793; F:defense/immunity protein activity; IEA.
DR InterPro: IPR003067; Crmsprzoite.
DR InterPro: IPR000884; TSPI.
DR Pfam: PF00090; tsp.1; 1.
DR PRINTS: PR01303; CRMSPRZOITE.
DR SMART: SM00209; TSPI; 1.
DR PROSITE: PS50092; TSPI; 1.
DR NON_TER 1
FT NON_TER 1
FT NON_TER 80
SQ SEQUENCE 80 AA; 9002 MW; 1CEEAE08E6C9E976 CRC64;

Query Match      85.3%; Score 81; DB 5; Length 80;
Best Local Similarity 89.5%; Pred. No. 3.5e-05;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 EKXIAMEKASSVFNVNS 20
DB 62 EKXICKMEKCSSVFNVNS 80

Search completed: March 10, 2004, 09:25:23
Job time : 20.7665 secs

```

```

ID Q9TVP9 PRELIMINARY; PRT; 80 AA.
AC Q9TVP9;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Circumsporozoite protein (Fragment).
GN CS.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MO, M6, M7, and M1;
RA "Sequence variation in the non-repeat region of the Plasmodium
RT falciparum glutamate rich protein (GLURP) from Brazil, Senegalese, and
RT Burmese field isolates and from laboratory strains.";
EL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ269955; CAB64242.1; -
DR EMBL; AJ269958; CAB64238.1; -
DR EMBL; AJ269959; CAB64239.1; -
DR EMBL; AJ269956; CAB64241.1; -
DR GO; GO:0003793; F:defense/immunity protein activity; IEA.
DR InterPro; IPR003067; Crcmsprzoite.
DR InterPro; IPR000884; TSP1.
DR Pfam; PF00090; tsp_1; 1.
DR PRINTS; PR01303; CRCMSPRZOITE.
DR SMART; SMO0209; TSP1; 1.
DR PROSITE; PS50092; TSP1; 1.
FT NON_TER 1
FT NON_TER 80
SQ SEQUENCE 80 AA; 9074 MW; A8F40C90DB1C033E CRC64;

Query Match 85.3%; Score 81; DB 5; Length 80;
Best Local Similarity 89.5%; Pred. No. 3.5e-05;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 EKKIAMEKASSVFNVNS 20
Db ||||| ||||| ||||| ||||| |||||
62 EKKICKMEKCSSVFNVNS 80

RESULT 10
Q9U0Q0 PRELIMINARY; PRT; 80 AA.
ID Q9U0Q0;
AC Q9U0Q0;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Circumsporozoite protein (Fragment).
GN CS.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=D28;
RA "Sequence variation in the non-repeat region of the Plasmodium
RT falciparum glutamate rich protein (GLURP) from Brazil, Senegalese, and
RT Burmese field isolates and from laboratory strains.";
EL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ269948; CAB64173.1; -
DR GO; GO:0003793; F:defense/immunity protein activity; IEA.
DR InterPro; IPR003067; Crcmsprzoite.
DR InterPro; IPR000884; TSP1.
DR Pfam; PF00090; tsp_1; 1.
DR PRINTS; PR01303; CRCMSPRZOITE.
DR SMART; SMO0209; TSP1; 1.
DR PROSITE; PS50092; TSP1; 1.
FT NON_TER 1
FT NON_TER 80
SQ SEQUENCE 80 AA; 9046 MW; BA7689D18F031C3E CRC64;

Query Match 85.3%; Score 81; DB 5; Length 80;
Best Local Similarity 89.5%; Pred. No. 3.5e-05;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 EKKIAMEKASSVFNVNS 20
Db ||||| ||||| ||||| ||||| |||||
62 EKKICKMEKCSSVFNVNS 80

RESULT 11
Q9U0P5 PRELIMINARY; PRT; 80 AA.
ID Q9U0P5;
AC Q9U0P5;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Circumsporozoite protein (Fragment).
GN CS.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=D4363;
RA "Sequence variation in the non-repeat region of the Plasmodium
RT falciparum glutamate rich protein (GLURP) from Brazil, Senegalese, and
RT Burmese field isolates and from laboratory strains.";
EL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ269965; CAB64184.1; -
DR GO; GO:0003793; F:defense/immunity protein activity; IEA.
DR InterPro; IPR003067; Crcmsprzoite.
DR InterPro; IPR000884; TSP1.
DR Pfam; PF00090; tsp_1; 1.
DR PRINTS; PR01303; CRCMSPRZOITE.
DR SMART; SMO0209; TSP1; 1.
DR PROSITE; PS50092; TSP1; 1.
FT NON_TER 1
FT NON_TER 80
SQ SEQUENCE 80 AA; 9000 MW; 03799BD18F0BF3B3 CRC64;

Query Match 85.3%; Score 81; DB 5; Length 80;
Best Local Similarity 89.5%; Pred. No. 3.5e-05;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 EKKIAMEKASSVFNVNS 20
Db ||||| ||||| ||||| ||||| |||||
62 EKKICKMEKCSSVFNVNS 80

RESULT 12
Q9TW76 PRELIMINARY; PRT; 80 AA.
ID Q9TW76;
AC Q9TW76;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Circumsporozoite protein (Fragment).
GN CS.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PALO ALTO, and D50;
RA "Sequence variation in the non-repeat region of the Plasmodium
RT falciparum glutamate rich protein (GLURP) from Brazil, Senegalese, and
RT Burmese field isolates and from laboratory strains.";
EL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ269954; CAB64179.1; -
DR EMBL; AJ269944; CAB64170.1; -

```

```

DR GO; GO:0003793; F:defense/immunity protein activity; IEA.
DR InterPro; IPR003067; Crmsprzoite.
DR Pfam; PF00090; tsp.1; 1.
DR PRINTS; PR01303; CRMSPRZOITE.
DR SMART; SM00209; TSP1; 1.
DR PROSITE; PS50092; TSP1; 1.
FT NON TER 1
FT NON TER 80
SQ SEQUENCE 80 AA; 9000 MW; 036C7E319APE13B3 CRC64;

Query Match      85.3%; Score 81; DB 5; Length 80;
Best Local Similarity 89.5%; Pred. No. 3.5e-05;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 EKKIARMEKASSVFNVVNS 20
   ||||| ||||| ||||| ||||| |||||
Db 62 EKKICKMEKCSSVFNVVNS 80

RESULT 6
Q9U0P1 PRELIMINARY; PRT; 80 AA.
AC Q9U0P1;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Circumsporozoite protein (Fragment).
GN CS.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=B1896;
RX MEDLINE=20542035; PubMed=11087922;
RA de Stricker K., Vuust J., Jepsen S., Oeuwray C., Theisen M.;
RT "Conservation and heterogeneity of the glutamate-rich protein (GLURP)
RT among field isolates and laboratory lines of Plasmodium falciparum."
RL Mol. Biochem. Parasitol. 111:123-130(2000).
DR ENBL; AJ269978; CAB64197.1; -.
DR GO; GO:0003793; F:defense/immunity protein activity; IEA.
DR InterPro; IPR003067; Crmsprzoite.
DR InterPro; IPR000884; TSP1.
DR Pfam; PF00090; tsp.1; 1.
DR PRINTS; PR01303; CRMSPRZOITE.
DR SMART; SM00209; TSP1; 1.
DR PROSITE; PS50092; TSP1; 1.
FT NON TER 1
FT NON TER 80
SQ SEQUENCE 80 AA; 9102 MW; BFC6C970CEFF0FA3E CRC64;

Query Match      85.3%; Score 81; DB 5; Length 80;
Best Local Similarity 89.5%; Pred. No. 3.5e-05;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 EKKIARMEKASSVFNVVNS 20
   ||||| ||||| ||||| ||||| |||||
Db 62 EKKICKMEKCSSVFNVVNS 80

RESULT 7
Q9TVQ0 PRELIMINARY; PRT; 80 AA.
AC Q9TVQ0;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Circumsporozoite protein (Fragment).
GN CS.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;

```

```

RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=D4393, and D4372;
RA de Stricker K., Vuust J., Jepsen S., Oeuwray C., Theisen M.;
RT "Sequence variation in the non-repeat region of the Plasmodium
RT falciparum glutamate rich protein (GLURP) from Brazil, Senegalese, and
RT Burmese field isolates and from laboratory strains."
RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
DR ENBL; AJ269966; CAB64185.1; -.
DR ENBL; AJ269964; CAB64183.1; -.
DR GO; GO:0003793; F:defense/immunity protein activity; IEA.
DR InterPro; IPR003067; Crmsprzoite.
DR InterPro; IPR000884; TSP1.
DR Pfam; PF00090; tsp.1; 1.
DR PRINTS; PR01303; CRMSPRZOITE.
DR SMART; SM00209; TSP1; 1.
DR PROSITE; PS50092; TSP1; 1.
FT NON TER 1
FT NON TER 80
SQ SEQUENCE 80 AA; 9059 MW; A756D1FFC41C1C21 CRC64;

Query Match      85.3%; Score 81; DB 5; Length 80;
Best Local Similarity 89.5%; Pred. No. 3.5e-05;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 EKKIARMEKASSVFNVVNS 20
   ||||| ||||| ||||| ||||| |||||
Db 62 EKKICKMEKCSSVFNVVNS 80

RESULT 8
Q9U0P9 PRELIMINARY; PRT; 80 AA.
AC Q9U0P9;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Circumsporozoite protein (Fragment).
GN CS.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=L1;
RA de Stricker K., Vuust J., Jepsen S., Oeuwray C., Theisen M.;
RT "Sequence variation in the non-repeat region of the Plasmodium
RT falciparum glutamate rich protein (GLURP) from Brazil, Senegalese, and
RT Burmese field isolates and from laboratory strains."
RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
DR ENBL; AJ269951; CAB64176.1; -.
DR GO; GO:0003793; F:defense/immunity protein activity; IEA.
DR InterPro; IPR003067; Crmsprzoite.
DR InterPro; IPR000884; TSP1.
DR Pfam; PF00090; tsp.1; 1.
DR PRINTS; PR01303; CRMSPRZOITE.
DR SMART; SM00209; TSP1; 1.
DR PROSITE; PS50092; TSP1; 1.
FT NON TER 1
FT NON TER 80
SQ SEQUENCE 80 AA; 9032 MW; ADED6F0E266AD98E CRC64;

Query Match      85.3%; Score 81; DB 5; Length 80;
Best Local Similarity 89.5%; Pred. No. 3.5e-05;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 EKKIARMEKASSVFNVVNS 20
   ||||| ||||| ||||| ||||| |||||
Db 62 EKKICKMEKCSSVFNVVNS 80

RESULT 9
Q9TVP9

```



```

RESULT 2
Q9UOP3      PRELIMINARY;      PRT;      80 AA.
ID Q9UOP3;
AC Q9UOP3;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Circumsporozoite protein (Fragment).
GN CS.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=D4405;
RA de Stricker K., Vuust J., Jepsen S., Oeuvray C., Theisen M.;
RT "Sequence variation in the non-repeat region of the Plasmodium
RT falciparum glutamate rich protein (GLURP) from Brazil, Senegalese, and
RT Burmese field isolates and from laboratory strains.";
RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ269969; CAB64188.1; -.
DR GO; GO:0003793; F:Defense/immunity protein activity; IEA.
DR InterPro; IPR003067; Crcmsprzoite.
DR Pfam; PF00090; tsp.1; 1.
DR PRINTS; PR01303; CRCMSPRZOITE.
DR SMART; SM00209; TSP1; 1.
DR PROSITE; PS50092; TSP1; 1.
FT NON TER 1
FT NON TER 80
SQ SEQUENCE 80 AA; 9033 MW; BA71EB0CD803193B CRC64;

Query Match      85.3%; Score 81; DB 5; Length 80;
Best Local Similarity 89.5%; Pred. No. 3.5e-05;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      2 EKXIAMEKASSVFNVVNS 20
      ||||| ||||| ||||| |||||
DB      62 EKXICKMEKCSSVFNVVNS 80

RESULT 3
Q9TW01      PRELIMINARY;      PRT;      80 AA.
ID Q9TW01;
AC Q9TW01;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Circumsporozoite protein (Fragment).
GN CS.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=WL1, and D7;
RA de Stricker K., Vuust J., Jepsen S., Oeuvray C., Theisen M.;
RT "Sequence variation in the non-repeat region of the Plasmodium
RT falciparum glutamate rich protein (GLURP) from Brazil, Senegalese, and
RT Burmese field isolates and from laboratory strains.";
RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ269960; CAB64240.1; -.
DR EMBL; AJ269964; CAB64172.1; -.
DR GO; GO:0003793; F:Defense/immunity protein activity; IEA.
DR InterPro; IPR003067; Crcmsprzoite.
DR Pfam; PF00090; tsp.1; 1.
DR PRINTS; PR01303; CRCMSPRZOITE.
DR SMART; SM00209; TSP1; 1.
DR PROSITE; PS50092; TSP1; 1.
FT NON TER 1
FT NON TER 80
SQ SEQUENCE 80 AA; 9123 MW; 4614EEC68F0B1434 CRC64;

Query Match      85.3%; Score 81; DB 5; Length 80;
Best Local Similarity 89.5%; Pred. No. 3.5e-05;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      2 EKXIAMEKASSVFNVVNS 20
      ||||| ||||| ||||| |||||
DB      62 EKXICKMEKCSSVFNVVNS 80

RESULT 5
Q9TW97      PRELIMINARY;      PRT;      80 AA.
ID Q9TW97;
AC Q9TW97;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Circumsporozoite protein (Fragment).
GN CS.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=D4403, and D4259;
RA de Stricker K., Vuust J., Jepsen S., Oeuvray C., Theisen M.;
RT "Sequence variation in the non-repeat region of the Plasmodium
RT falciparum glutamate rich protein (GLURP) from Brazil, Senegalese, and
RT Burmese field isolates and from laboratory strains.";
RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ269968; CAB64187.1; -.
DR EMBL; AJ269962; CAB64181.1; -.

```

```

Query Match      85.3%; Score 81; DB 5; Length 80;
Best Local Similarity 89.5%; Pred. No. 3.5e-05;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      2 EKXIAMEKASSVFNVVNS 20
      ||||| ||||| ||||| |||||
DB      62 EKXICKMEKCSSVFNVVNS 80

RESULT 4
Q9UOP7      PRELIMINARY;      PRT;      80 AA.
ID Q9UOP7;
AC Q9UOP7;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Circumsporozoite protein (Fragment).
GN CS.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=D4230;
RA de Stricker K., Vuust J., Jepsen S., Oeuvray C., Theisen M.;
RT "Sequence variation in the non-repeat region of the Plasmodium
RT falciparum glutamate rich protein (GLURP) from Brazil, Senegalese, and
RT Burmese field isolates and from laboratory strains.";
RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ269961; CAB64180.1; -.
DR GO; GO:0003793; F:Defense/immunity protein activity; IEA.
DR InterPro; IPR003067; Crcmsprzoite.
DR Pfam; PF00090; tsp.1; 1.
DR PRINTS; PR01303; CRCMSPRZOITE.
DR SMART; SM00209; TSP1; 1.
DR PROSITE; PS50092; TSP1; 1.
FT NON TER 1
FT NON TER 80
SQ SEQUENCE 80 AA; 9047 MW; BA769C90DB031C3E CRC64;

Query Match      85.3%; Score 81; DB 5; Length 80;
Best Local Similarity 89.5%; Pred. No. 3.5e-05;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      2 EKXIAMEKASSVFNVVNS 20
      ||||| ||||| ||||| |||||
DB      62 EKXICKMEKCSSVFNVVNS 80

RESULT 5
Q9TW97      PRELIMINARY;      PRT;      80 AA.
ID Q9TW97;
AC Q9TW97;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Circumsporozoite protein (Fragment).
GN CS.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=D4403, and D4259;
RA de Stricker K., Vuust J., Jepsen S., Oeuvray C., Theisen M.;
RT "Sequence variation in the non-repeat region of the Plasmodium
RT falciparum glutamate rich protein (GLURP) from Brazil, Senegalese, and
RT Burmese field isolates and from laboratory strains.";
RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ269968; CAB64187.1; -.
DR EMBL; AJ269962; CAB64181.1; -.

```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 10, 2004, 08:58:54 ; Search time 19.7665 Seconds
(without alignments)
319.245 Million cell updates/sec

Title: US-09-848-834A-3

Perfect score: 95
Sequence: 1 DEKKIAKMEKASSVFNVVNS 20

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL 25:
1: sp_archaea:
2: sp_bacteria:
3: sp_fungi:
4: sp_human:
5: sp_invertebrate:
6: sp_mammal:
7: sp_mhc:
8: sp_organelle:
9: sp_phase:
10: sp_plant:
11: sp_rodent:
12: sp_virus:
13: sp_vertebrate:
14: sp_unclassified:
15: sp_rvirus:
16: sp_bacteriap:
17: sp_archaeap:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	81	85.3	69	5 Q8MZK7	Q8MZK7 plasmodium
2	81	85.3	80	5 Q9U0P3	Q9U0P3 plasmodium
3	81	85.3	80	5 Q9TW01	Q9TW01 plasmodium
4	81	85.3	80	5 Q9U0P7	Q9U0P7 plasmodium
5	81	85.3	80	5 Q9TW57	Q9TW57 plasmodium
6	81	85.3	80	5 Q9U0E1	Q9U0E1 plasmodium
7	81	85.3	80	5 Q9TVQ0	Q9TVQ0 plasmodium
8	81	85.3	80	5 Q9U0P9	Q9U0P9 plasmodium
9	81	85.3	80	5 Q9TVP9	Q9TVP9 plasmodium
10	81	85.3	80	5 Q9U0C0	Q9U0C0 plasmodium
11	81	85.3	80	5 Q9U0P5	Q9U0P5 plasmodium
12	81	85.3	80	5 Q9TW76	Q9TW76 plasmodium
13	81	85.3	80	5 Q9U0U4	Q9U0U4 plasmodium
14	81	85.3	80	5 Q9TW00	Q9TW00 plasmodium
15	81	85.3	80	5 Q9U0P6	Q9U0P6 plasmodium
16	81	85.3	80	5 Q9U0P2	Q9U0P2 plasmodium

17	81	85.3	80	5 Q9U0P8	Q9U0P8 plasmodium
18	81	85.3	80	5 Q9U0P4	Q9U0P4 plasmodium
19	81	85.3	80	5 Q9TVN9	Q9TVN9 plasmodium
20	81	85.3	80	5 Q9TW83	Q9TW83 plasmodium
21	81	85.3	115	5 Q25837	Q25837 plasmodium
22	81	85.3	115	5 Q25839	Q25839 plasmodium
23	81	85.3	115	5 Q9U3J4	Q9U3J4 plasmodium
24	81	85.3	115	5 Q25835	Q25835 plasmodium
25	81	85.3	115	5 Q25836	Q25836 plasmodium
26	81	85.3	115	5 Q9TVW7	Q9TVW7 plasmodium
27	81	85.3	117	5 Q25795	Q25795 plasmodium
28	81	85.3	117	5 Q25796	Q25796 plasmodium
29	81	85.3	117	5 Q25797	Q25797 plasmodium
30	81	85.3	117	5 Q25794	Q25794 plasmodium
31	81	85.3	356	5 Q819G5	Q819G5 plasmodium
32	81	85.3	360	5 Q819H7	Q819H7 plasmodium
33	81	85.3	360	5 Q819H6	Q819H6 plasmodium
34	81	85.3	363	5 Q819H3	Q819H3 plasmodium
35	81	85.3	364	5 Q819C4	Q819C4 plasmodium
36	81	85.3	368	5 Q819C3	Q819C3 plasmodium
37	81	85.3	380	5 Q819C8	Q819C8 plasmodium
38	81	85.3	380	5 Q819C1	Q819C1 plasmodium
39	81	85.3	385	5 Q819I9	Q819I9 plasmodium
40	81	85.3	385	5 Q819I8	Q819I8 plasmodium
41	81	85.3	388	5 Q819H4	Q819H4 plasmodium
42	81	85.3	389	5 Q819U2	Q819U2 plasmodium
43	81	85.3	389	5 Q81056	Q81056 plasmodium
44	81	85.3	392	5 Q819H8	Q819H8 plasmodium
45	81	85.3	392	5 Q819H1	Q819H1 plasmodium

ALIGNMENTS

RESULT 1

Q8MZK7 Q8MZK7 PRELIMINARY; PRT; 69 AA.
AC Q8MZK7;
DT 01-OCT-2002 (TremBLrel. 22, Created)
DT 01-OCT-2002 (TremBLrel. 22, Last sequence update)
DT 01-JUN-2003 (TremBLrel. 24, Last annotation update)
DE Circumsporozoite protein II (Fragment).
GN CSPII.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=FCCL/HN;
RA Chen H.H., Yu X.B., Wu Z.D., Xu J.;
RT "Gene cloning of circumsporozoite protein (CSP) II gene from Plasmodium falciparum (FCCL/HN).";
RL Submitted (MAR-2002) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AY093672; AAM19072.1;
DR GO; GO:0003793; F:defense/immunity protein activity; IEA.
DR InterPro; IPR003067; Crcmsprzoite.
DR InterPro; IPR000884; TSPI.
DR Pfam; PF00090; tsp; 1.
DR PRINTS; PR01303; CRCMSPRZOITE.
DR SMART; SM00209; TSPI; 1.
DR PROSITE; PS50092; TSPI; 1.
FT NON_TER
SQ SEQUENCE 69 AA; 7670 MW; F3E1AAS4671FE6FF CRC64;

Query Match 85.3%; Score 81; DB 5; Length 69;
Best Local Similarity 89.5%; Pred. No. 3e-05;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 EKXIAKMEKASSVFNVVNS 20

DB 37 EKXIAKMEKASSVFNVVNS 55

QY 2 HWSYGLRP 9
|||||
Db 23 HWSFGLSP 30

RESULT 14

A21114
gonadoliberin - chum salmon
C;Species: Oncorhynchus keta (chum salmon)
C;Date: 10-Aug-1990 #sequence_revision 10-Aug-1990 #text_change 18-Jun-1993
C;Accession: A21114
R;Sherwood, N.; Biden, L.; Brownstein, M.; Spiess, J.; Rivier, J.; Vale, W.
Proc. Natl. Acad. Sci. U.S.A. 80, 2794-2798, 1983
A;Title: Characterization of a teleost gonadotropin-releasing hormone.
A;Reference number: A21114; MUID:83195140; PMID:6341999
A;Accession: A21114
A;Status: preliminary
A;Molecule type: protein
A;Residues: 1-10 <SH>

Query Match 72.2%; Score 39; DB 2; Length 10;
Best Local Similarity 75.0%; Pred. No. 0.33;
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 HWSYGLRP 9
|||||
Db 2 HWSYGLWP 9

RESULT 15

I51092
gonadotropin releasing hormone - chinook salmon (fragment)
C;Species: Oncorhynchus tshawytscha (chinook salmon)
C;Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 09-Aug-1997
C;Accession: I51092
R;Klungland, H.; Lorens, J.B.; Andersen, O.; Kisen, G.O.; Alestrom, P.
Mol. Cell. Endocrinol. 84, 167-174, 1992
A;Title: The Atlantic salmon prepro-gonadotropin releasing hormone gene and mRNA.
A;Reference number: I51040; MUID:92267241; PMID:1587389
A;Accession: I51092
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-74 <KL>
A;Cross-references: EMBL:X79711; NID:G499322; PID:G499323
C;Genetics:
A;Gene: GnRH
A;Introns: 38/3; 65/3

Query Match 72.2%; Score 39; DB 2; Length 74;
Best Local Similarity 75.0%; Pred. No. 2.9;
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 HWSYGLRP 9
|||||
Db 17 HWSYGLWP 24

Search completed: March 10, 2004, 09:16:38
Job time : 4.17166 secs

```

Best Local Similarity 87.5%; Pred. No. 0.088;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 HWSYGLRP 9
   |||||:|
Db 25 HWSYGLQP 32

RESULT 10
150739
gonadotropin-releasing hormone - Cichlid (Haplochromis burtoni)
C:Species: Haplochromis burtoni
C>Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 21-Jul-2000
C:Accession: I50739
R;White, S.A.; Kaaten, T.L.; Bond, C.T.; Adelman, J.P.; Pernald, R.D.
Proc. Natl. Acad. Sci. U.S.A. 92, 8363-8367, 1995
A:Title: Three gonadotropin-releasing hormone genes in one organism suggest novel roles
A:Reference number: I50739; MUID:95396797; PMID:7667296
A:Accession: I50739
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: mRNA
A:Residues: 1-98 <WHI>
A:Cross-references: EMBL:U91865; NID:G905398; PIDN:AA053691.1; PID:G905399
C:Superfamily: gonadoliberin

Query Match 85.2%; Score 46; DB 2; Length 98;
Best Local Similarity 87.5%; Pred. No. 0.22; Indels 0; Gaps 0;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 HWSYGLRP 9
   |||||:|
Db 24 HWSYGLSP 31

RESULT 11
E95361
probable muconate cycloisomerase (EC 5.5.1.1) [imported] - Sinorhizobium meliloti (strain
C:Species: Sinorhizobium meliloti
C>Date: 24-Aug-2001 #sequence_revision 24-Aug-2001 #text_change 30-Sep-2001
C:Accession: E95361
R;Barnett, M.J.; Fisher, R.F.; Jones, T.; Komp, C.; Abola, A.P.; Barloy-Hubler, F.; Bows
L.; Kalmann, S.; Keating, D.H.; Palm, C.; Peck, M.C.; Surzycki, R.; Wells, D.H.; Yeh, K.C.
Proc. Natl. Acad. Sci. U.S.A. 98, 9893-9898, 2001
A:Title: Nucleotide sequence and predicted functions of the entire Sinorhizobium meliloti
A:Reference number: A95262; MUID:21396509; PMID:11481432
A:Accession: E95361
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-374 <KUR>
A:Cross-references: GB:AB006469; PIDN:AAK65455.1; PID:G14523923; GSPDB:GNO00165
A:Experimental source: strain 1021, megaplasmid pSYMA
R;Galibert, F.; Finan, T.M.; Long, S.R.; Puhler, A.; Abola, P.; Ampe, P.; Barloy-Hubler,
pela, D.; Chain, P.; Cowie, A.; Davis, R.W.; Dreano, S.; Federspiel, N.A.; Fisher, R.F.;
L.; Hyman, R.W.; Jones, T.
Science 293, 668-672, 2001
A:Authors: Kahn, D.; Kahn, M.L.; Kalmann, S.; Keating, D.H.; Kiss, E.; Komp, C.; Lelaure,
Rebault, P.; Vandenbol, M.; Vorholter, F.J.; Weidner, S.; Wells, D.H.; Wong, K.; Yeh, K.
A:Title: The composite genome of the legume symbiont Sinorhizobium meliloti.
A:Reference number: A96039; MUID:21368234; PMID:11474104
A:Contents: annotation
C:Genetics:
A:Gene: Sma1461
A:Genome: plasmid
C:Keywords: intramolecular lyase; isomerase

Query Match 79.6%; Score 43; DB 2; Length 374;
Best Local Similarity 85.7%; Pred. No. 3.3; Indels 0; Gaps 0;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 HWSYGLR 8
   |||||:|
Db 21 HWSYGLR 27

```

```

RESULT 12
RHID15
gonadoliberin I precursor - sharptooth catfish
N;Alternate names: gonadoliberin, catfish-type; gonadotropin-releasing hormone I (GnRH-
N;Contains: gonadoliberin I; gonadoliberin I-associated protein form I; gonadoliberin I
C:Species: Clarias gariepinus (sharptooth catfish)
C>Date: 30-Sep-1993 #sequence_revision 18-Mar-1997 #text_change 18-Jun-1999
C:Accession: S45602; S45601; JC1442; S42936; S42937; H.
R;Bogerd, J.; Zandbergen, T.; Andersson, E.; Goos, H.
Eur. J. Biochem. 222, 541-549, 1994
A:Title: Isolation, characterization and expression of cDNAs encoding the catfish-type
A:Reference number: S45600; MUID:94291651; PMID:8020492
A:Accession: S45602
A:Molecule type: mRNA
A:Residues: 1-80 <BOG1>
A:Cross-references: EMBL:X78049; NID:G459433; PIDN:CAA54971.1; PID:G459434
A:Note: gonadoliberin I-associated protein form I
A:Accession: S45601
A:Molecule type: mRNA
A:Residues: 1-46, 'S', 48-59, 'G', 61-80 <BOG2>
A:Cross-references: EMBL:X78048; NID:G459431; PIDN:CAA54970.1; PID:G459432
A:Note: gonadoliberin I-associated protein form II, presumed to be a polymorphic form
R;Bogerd, J.; Li, K.W.; Janssen-Dommersholt, C.; Goos, H.
Biochem. Biophys. Res. Commun. 187, 127-134, 1992
A:Title: Two gonadotropin-releasing hormones from African catfish (Clarias gariepinus).
A:Reference number: JC1242; MUID:92392313; PMID:1520292
A:Accession: JC1242
A:Molecule type: protein
A:Residues: 22-31 <BOG3>
A:Experimental source: brain
C:Superfamily: gonadoliberin
C:Keywords: amidated carboxyl end; hormone; hypothalamus; pyroglutamic acid
F:1-21/Domain: signal sequence #status predicted <sig>
F:22-31/Product: gonadoliberin I #status experimental <MAT1>
F:35-80/Product: gonadoliberin I-associated protein #status predicted <MAT2>
F:22/Modified site: pyrrolidone carboxylic acid (Gln) (in mature form) #status experime
F:31/Modified site: amidated carboxyl end (Gly) (amide in mature form from following gl

Query Match 77.8%; Score 42; DB 1; Length 80;
Best Local Similarity 75.0%; Pred. No. 0.92; Indels 1; Gaps 0;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 HWSYGLRP 9
   |||||:|
Db 23 HWSYGLNP 30

RESULT 13
JC7393
medaka-type gonadotropin-releasing hormone precursor - Japanese medaka
C:Species: Oryzias latipes (Japanese medaka)
C>Date: 17-Nov-2000 #sequence_revision 17-Nov-2000 #text_change 20-Jan-2003
C:Accession: JC7393
R;Okubo, K.; Amano, M.; Yoshiura, Y.; Suetake, H.; Aida, K.
Biochem. Biophys. Res. Commun. 276, 298-303, 2000
A:Title: A novel form of gonadotropin-releasing hormone in the medaka, Oryzias latipes.
A:Reference number: JC7393
A:Contents: Brain
A:Accession: JC7393
A:Molecule type: mRNA
A:Residues: 1-91 <OKU>
A:Cross-references: DBJ:AB041333
C:Comment: This protein plays the roles as a hypophysiotropic factor, and a physiologic
C:Genetics:
A:Gene: mdgnrh
C:Superfamily: gonadoliberin
C:Keywords: Brain

Query Match 77.8%; Score 42; DB 2; Length 91;
Best Local Similarity 75.0%; Pred. No. 1.1; Indels 1; Gaps 0;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

```

R.Tan, L.; Rousseau, P.
 Biochem. Biophys. Res. Commun. 109, 1061-1071, 1982
 A:Title: The chemical identity of the immunoreactive LHRH-like peptide biosynthesized in
 A:Reference number: A90108; MUID:83126573; PMID:6760865
 A:Accession: A90108
 A:Molecule type: Protein
 A:Residues: 24-33 <TRAN>
 A:Experimental source: placental trophoblasts
 R:Leibovitz, D.; Koch, Y.; Pitzer, F.; Fridkin, M.; Dantes, A.; Baumeister, W.; Amsterda
 FEBS Lett. 346, 203-206, 1994
 A:Title: Sequential degradation of the neuropeptide gonadotropin-releasing hormone by th
 A:Reference number: S45718; MUID:94283597; PMID:8013634
 A:Contents: annotation; degradation pathway of synthetic hormone
 C:Genetics:
 A:Gene: GDB:GNRH; LHRH; GRH
 A:Cross-references: GDB:133746; OMIM:227200; OMIM:152760
 A:Map position: 8p21-8p11.2
 A:Introns: 47/3; 79/3
 C:Function:
 A:Description: gonadoliberein stimulates pituitary secretion of lutropin and follitropin
 A:Note: gonadoliberein-associated protein may have prolactin release inhibiting activity
 C:Superfamily: gonadoliberein
 C:Keywords: amidated carboxyl end; hormone; hypothalamus; placenta; pyroglutamic acid
 F:1-23/Domain: signal sequence #status predicted <SIG>
 F:24-92/Product: progadoliberein #status predicted <PGN>
 F:24-33/Product: gonadoliberein #status experimental <MAT>
 F:37-92/Product: gonadoliberein-associated protein #status predicted <GAP>
 F:37-92/Product: prolactin release-inhibiting factor #status predicted <PIF>
 F:24/Modified site: pyroglutamic acid (Gln) (in mature form) #status experimental
 F:33/Modified site: amidated carboxyl end (Gly) (amide in mature form from following gly

Query Match 96.3%; Score 52; DB 1; Length 92;
 Best Local Similarity 100.0%; Pred. No. 0.017;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 A:Introns: 47/3; 79/3

QY 2 HWSYGLRP 9
 DB 25 HWSYGLRP 32

RESULT 7
 RHRTG
 Gonadoliberein precursor - rat
 N:Alternate names: gonadoliberein-associated protein (GAP); gonadotropin releasing hormon
 N:Contains: gonadoliberein; prolactin release-inhibiting factor
 C:Species: Rattus norvegicus (Norway rat)
 C:Date: 31-Mar-1988 #sequence revision 31-Mar-1988 #text_change 18-Jun-1999
 C:Accession: A40147; B26173; A48410
 R:Bond, C.T.; Hayflick, J.S.; Seeburg, P.H.; Adelman, J.P.
 Mol. Endocrinol. 3, 1257-1262, 1989
 A:Title: The rat gonadotropin-releasing hormone: SH locus: structure and hypothalamic ex
 A:Reference number: A40147; MUID:89384661; PMID:2476669
 A:Accession: A40147
 A:Molecule type: DNA
 A:Residues: 1-92 <BON>
 A:Cross-references: GB:M31670; NID:G204447; PIDN:AAA41264.1; PID:G204448
 R:Adelman, J.P.; Mason, A.J.; Hayflick, J.S.; Seeburg, P.H.
 Proc. Natl. Acad. Sci. U.S.A. 83, 179-183, 1986
 A:Title: Isolation of the gene and hypothalamic cDNA for the common precursor of gonadob
 A:Reference number: A94090; MUID:86094338; PMID:2867548
 A:Accession: B26173
 A:Molecule type: mRNA
 A:Residues: 1-92 <AD>
 A:Cross-references: GB:M12579; NID:G204445; PIDN:AAA41263.1; PID:G204446
 R:Maier, C.C.; Marchetti, B.; LeBoeuf, R.D.; Blalock, J.E.
 Cell. Mol. Neurobiol. 12, 447-454, 1992
 A:Title: Thymocytes express a mRNA that is identical to hypothalamic luteinizing hormone
 A:Reference number: A48410; MUID:93105480; PMID:1468115
 A:Accession: A48410
 A:Status: preliminary
 A:Molecule type: mRNA
 A:Cross-references: GB:S50870; NID:G262059; PIDN:AAB24572.1; PID:G262060
 A:Experimental source: thymus

A:Note: sequence extracted from NCBI backbone (NCBIN:121082, NCSIP:121083)
 C:Genetics:
 A:Introns: 47/3; 79/3
 C:Function:
 A:Description: stimulates pituitary secretion of lutropin and follitropin
 A:Note: gonadoliberein-associated protein may have prolactin release inhibiting activity
 C:Superfamily: gonadoliberein
 C:Keywords: amidated carboxyl end; hormone; hypothalamus; placenta; pyroglutamic acid;
 F:1-23/Domain: signal sequence #status predicted <SIG>
 F:24-92/Product: progadoliberein #status predicted <PGN>
 F:24-33/Product: gonadoliberein #status predicted <GDM>
 F:37-92/Product: prolactin release-inhibiting factor #status predicted <PIF>
 F:24/Modified site: pyroglutamic acid (Gln) (in mature form) #status predicted
 F:33/Modified site: amidated carboxyl end (Gly) (amide in mature form from following gl

Query Match 96.3%; Score 52; DB 1; Length 92;
 Best Local Similarity 100.0%; Pred. No. 0.017;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 HWSYGLRP 9
 DB 25 HWSYGLRP 32

RESULT 8

RHAQI
 gonadoliberein I - American alligator
 N:Alternate names: gonadotropin-releasing hormone I
 C:Species: Alligator mississippiensis (American alligator)
 C:Date: 31-Mar-1993 #sequence_revision 31-Mar-1993 #text_change 18-Mar-1997
 C:Accession: A60066
 R:Lovejoy, D.A.; Fischer, W.H.; Parker, D.B.; McRory, J.E.; Park, M.; Lance, V.; Swans
 Regul. Pept. 33, 105-116, 1991
 A:Title: Primary structure of two forms of gonadotropin-releasing hormone from brains
 A:Reference number: A60066; MUID:91352338; PMID:1882082
 A:Accession: A60066
 A:Molecule type: Protein
 A:Residues: 1-10 <LOW>
 C:Superfamily: gonadoliberein
 C:Keywords: amidated carboxyl end; hormone; hypothalamus; pyroglutamic acid
 F:1/Modified site: pyroglutamic acid (Gln) #status experimental
 F:10/Modified site: amidated carboxyl end (Gly) #status experimental

Query Match 88.9%; Score 48; DB 1; Length 10;
 Best Local Similarity 87.5%; Pred. No. 0.0077;
 Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 HWSYGLRP 9
 DB 2 HWSYGLRP 9

RESULT 9

IS0644
 gonadoliberein I precursor - chicken
 N:Alternate names: gonadotropin-releasing hormone I
 C:Species: Gallus gallus (Chicken)
 C:Date: 21-Feb-1997 #sequence_revision 21-Feb-1997 #text_change 16-Jul-1999
 C:Accession: IS0644; S33507
 R:Dunn, I.C.; Chen, Y.; Hook, C.; Sharp, P.J.; Sang, H.M.
 J. Mol. Endocrinol. 11, 19-29, 1993
 A:Title: Characterization of the chicken preprogonadotropin-releasing hormone-I gene.
 A:Reference number: IS0644; MUID:94059355; PMID:7902095
 A:Accession: IS0644
 A:Status: translated from GB/ENBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-92 <DU2>
 A:Cross-references: EMBL:X69491; NID:9496326; PIDN:CAA49246.1; PID:G311612
 C:Genetics:
 A:Introns: 47/3; 79/3
 C:Superfamily: gonadoliberein

Query Match 88.9%; Score 48; DB 2; Length 92;

A;Residues: 1-10 <BUR>
 A;Note: the natural and synthetic hormones have the same biological activity
 C;Comment: This hypothalamic hormone stimulates the secretion of both luteinizing and follicle stimulating hormone
 C;Superfamily: gonadoliberin
 C;Keywords: amidated carboxyl end; hormone; hypothalamus; pyroglutamic acid
 F;1/Modified site: pyroglutamate carboxylic acid (Gln) #status experimental
 F;10/Modified site: amidated carboxyl end (Gly) #status experimental

Query Match 96.3%; Score 52; DB 1; Length 10;
 Best Local Similarity 100.0%; Pred. No. 0.0015;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 HWSYGLRP 9
 |||||
 DB 2 HWSYGLRP 9
 |||||

RESULT 3
 I78541
 gonadoliberin precursor - rhesus macaque (fragment)
 N;Alternate names: luteinizing hormone releasing hormone
 C;Species: Macaca mulatta (rhesus macaque)
 C;Date: 02-Aug-1996 #sequence_revision 02-Aug-1996 #text_change 16-Jul-1999
 C;Accession: I78541
 R;Ma, Y.J.; Costa, M.E.; Ojeda, S.R.
 Neuroendocrinology 60, 346-359, 1994
 A;Title: Developmental expression of the genes encoding transforming growth factor alpha
 A;Reference number: I58134; MUID:95124501; PMID:7545971
 A;Accession: I78541
 A;Status: preliminary; translated from GB/EMBL/DBJ
 A;Molecule type: mRNA
 A;Residues: 1-67 <RES>
 A;Cross-references: GB:S75918; NID:G912831; PIDN:AA33096.1; PID:G912832
 C;Superfamily: gonadoliberin

Query Match 96.3%; Score 52; DB 2; Length 67;
 Best Local Similarity 100.0%; Pred. No. 0.012;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 HWSYGLRP 9
 |||||
 DB 7 HWSYGLRP 14
 |||||

RESULT 4
 I51423
 gonadoliberin precursor - African clawed frog
 N;Alternate names: luteinizing hormone releasing hormone
 C;Species: Xenopus laevis (African clawed frog)
 C;Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 16-Jul-1999
 C;Accession: I51423
 R;Hayes, W.P.; Wray, S.; Battey, J.F.
 Endocrinology 134, 1835-1845, 1994
 A;Title: The frog GnRH-I gene has a mammalian-like expression pattern and conserved domain
 A;Reference number: I51423; MUID:94185563; PMID:8137750
 A;Accession: I51423
 A;Status: preliminary; translated from GB/EMBL/DBJ
 A;Molecule type: DNA
 A;Residues: 1-89 <HAY>
 A;Cross-references: GB:L28040; NID:G496291; PIDN:AAA49728.1; PID:G496292
 C;Genetics:
 A;Gene: GnRH-I
 C;Superfamily: gonadoliberin

Query Match 96.3%; Score 52; DB 2; Length 89;
 Best Local Similarity 100.0%; Pred. No. 0.016;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 HWSYGLRP 9
 |||||
 DB 25 HWSYGLRP 32
 |||||

RESULT 5

RHMSG
 gonadoliberin precursor - mouse
 N;Alternate names: gonadotropin-releasing hormone (GnRH); luteinizing hormone releasing hormone
 C;Contains: gonadoliberin; gonadoliberin-associated protein (GAP)
 C;Species: Mus musculus (house mouse)
 C;Date: 31-Dec-1993 #sequence_revision 18-Mar-1997 #text_change 18-Jun-1999
 C;Accession: A47578
 R;Mason, A.J.; Hayflick, J.S.; Zoeller, R.T.; Young III, W.S.; Phillips, H.S.; Nikolic
 Science 234, 1366-1371, 1986
 A;Title: A deletion truncating the gonadotropin-releasing hormone gene is responsible for
 A;Reference number: A47578; MUID:87069928; PMID:3024317
 A;Accession: A47578
 A;Molecule type: DNA
 A;Residues: 1-90 <MAS>
 A;Cross-references: EMBL:M14872; NID:G193576; PIDN:AAA37717.1; PID:G387175
 C;Genetics:
 A;Introns: 45/3; 77/3
 C;Function:
 A;Description: gonadoliberin stimulates pituitary secretion of lutropin and follitropin
 N;Alternate names: gonadoliberin-associated protein may have prolactin release inhibiting activity
 C;Superfamily: gonadoliberin
 C;Keywords: amidated carboxyl end; hormone; hypothalamus; pyroglutamic acid
 F;1-23/Domain: signal sequence #status predicted <SIG>
 F;22-31/Product: gonadoliberin #status predicted <GLB>
 F;35-90/Product: gonadoliberin-associated protein #status predicted <GAP>
 F;22/Modified site: pyroglutamate carboxylic acid (Gln) (in mature form) #status predicted
 F;31/Modified site: amidated carboxyl end (Gly) (amide in mature form from following 9)

Query Match 96.3%; Score 52; DB 1; Length 90;
 Best Local Similarity 100.0%; Pred. No. 0.016;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 HWSYGLRP 9
 |||||
 DB 23 HWSYGLRP 30
 |||||

RESULT 6

RHUG
 gonadoliberin precursor [validated] - human
 N;Alternate names: gonadotropin releasing hormone (GnRH); luteinizing hormone releasing hormone
 C;Contains: gonadoliberin-associated protein (GAP); progadoliberin
 C;Species: Homo sapiens (man)
 C;Date: 17-Mar-1987 #sequence_revision 21-Jul-1995 #text_change 08-Dec-2000
 C;Accession: S05308; A26173; A93342; A90108; A01410; S45718
 R;Hayflick, J.S.; Adelman, J.P.; Seeburg, P.H.
 Nucleic Acids Res. 17, 6403-6404, 1989
 A;Title: The complete nucleotide sequence of the human gonadotropin-releasing hormone
 A;Reference number: S05308; MUID:89366682; PMID:2671939
 A;Accession: S05308
 A;Status: translation not shown
 A;Molecule type: DNA
 A;Residues: 1-92 <HAY>
 A;Cross-references: EMBL:X15215; NID:G31955; PIDN:CAA33285.1; PID:G31956
 R;Adelman, J.P.; Mason, A.J.; Hayflick, J.S.; Seeburg, P.H.
 Proc. Natl. Acad. Sci. U.S.A. 83, 179-183, 1986
 A;Title: Isolation of the gene and hypothalamic cDNA for the common precursor of gonadotropin-releasing hormone and growth hormone releasing hormone
 A;Reference number: A94090; MUID:86094338; PMID:2867548
 A;Accession: A26173
 A;Molecule type: mRNA
 A;Residues: 1-92 <ADE>
 A;Cross-references: GB:M12578; NID:G183418; PIDN:AAA35916.1; PID:G386749
 A;Experimental source: hypothalamus
 R;Seeburg, P.H.; Adelman, J.P.
 Nature 311, 666-668, 1984
 A;Title: Characterization of cDNA for precursor of human luteinizing hormone releasing hormone
 A;Reference number: A93342; MUID:85012739; PMID:6090951
 A;Accession: A93342
 A;Molecule type: mRNA
 A;Residues: 1-15, 'S', 17-92 <SEE>
 A;Cross-references: GB:X01059; NID:G34356; PIDN:CAA35526.1; PID:G34357
 A;Experimental source: placenta

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 10, 2004, 08:58:54 ; Search time 3.11284 Seconds
(without alignments)
309.015 Million cell updates/sec

Title: US-09-848-834A-1

Perfect score: 54

Sequence: 1 XHWSYGLRPX 10

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR 78:*

1: PIR1:*

2: PIR2:*

3: PIR3:*

4: PIR4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	52	96.3	10	1 RHPOG	gonadoliberin - pi
2	52	96.3	10	1 RHSHG	gonadoliberin - sh
3	52	96.3	67	2 I78541	gonadoliberin prec
4	52	96.3	89	2 I51423	gonadoliberin prec
5	52	96.3	90	1 RHMSG	gonadoliberin prec
6	52	96.3	92	1 RHUUG	gonadoliberin prec
7	52	96.3	92	1 RHRTG	gonadoliberin prec
8	48	88.9	10	1 RHAQ1	gonadoliberin I -
9	48	88.9	92	2 I50644	gonadoliberin I pr
10	46	85.2	98	2 I50739	gonadotropin-relea
11	43	79.6	374	2 E95361	probable muconate
12	42	77.8	80	1 RHID1S	gonadoliberin I pr
13	42	77.8	91	2 JC7393	medaka-type gonad
14	39	72.2	10	2 A21114	gonadoliberin - ch
15	39	72.2	74	2 I51092	gonadotropin relea
16	39	72.2	82	2 I51180	gonadotropin-relea
17	39	72.2	82	2 I51355	gonadotropin relea
18	39	72.2	82	2 I51365	gonadotropin relea
19	39	72.2	82	2 I51331	gonadotropin relea
20	39	72.2	90	2 JC7395	salmon-type gonad
21	39	72.2	90	2 A23735	gonadoliberin prec
22	39	72.2	90	2 I51095	gonadoliberin prec
23	39	72.2	388	2 C72710	probable fmu prote
24	39	72.2	1444	2 T18856	angiogenesis inhib
25	38	70.4	161	2 D84472	hypothetical prote
26	38	70.4	293	2 G72699	hypothetical prote
27	38	70.4	501	2 T32848	hypothetical prote
28	37	68.5	270	2 F88035	protein M01D1.7 [l
29	37	68.5	345	2 A58519	hypothetical 345 p

ALIGNMENTS

RESULT 1

RHPOG

gonadoliberin - pig
C:Species: Sus scrofa domestica (domestic pig)

C>Date: 13-Jul-1981 #sequence_revision 13-Jul-1981 #text_change 18-Mar-1997

C:Accession: A01411

R:Baba, Y.; Matsuo, H.; Schally, A.V.

Biochem. Biophys. Res. Commun. 44, 459-463, 1971

A>Title: Structure of the porcine LH and FSH-releasing hormone. II. Confirmation of th

A:Reference number: A90176; MUID:72114303; PMID:4946067

A:Accession: A01411

A:Molecule type: protein

A:Residues: 1-10 <BAB>

R:Matsuo, H.; Arimura, A.; Nair, R.M.G.; Schally, A.V.

Biochem. Biophys. Res. Commun. 45, 822-827, 1971

A>Title: Synthesis of the porcine LH and FSH-releasing hormone by the solid-phase meth

A:Reference number: A90176; MUID:72055376; PMID:4942726

A:Contents: annotation; synthesis

A>Note: the synthetic and natural hormones have the same physicochemical and biological

R:Baba, Y.; Arimura, A.; Schally, A.V.

Biochem. Biophys. Res. Commun. 45, 483-487, 1971

A>Title: On the tryptophan residue in porcine LH and FSH-releasing hormones.

A:Reference number: A90175; MUID:72117544; PMID:4946275

A:Contents: annotation

A>Note: Trp-3 appears to be essential for biological activity

C:Comment: This hypothalamic hormone stimulates the secretion of both luteinizing and l

C:Superfamily: gonadoliberin

C:Keywords: amidated carboxyl end; hormone; hypothalamus; pyroglutamic acid

F:1/Modified site: pyrrolidone carboxylic acid (Gln) #status experimental

F:10/Modified site: amidated carboxyl end (Gly) #status experimental

Query Match 96.3%; Score 52; DB 1; Length 10;

Best Local Similarity 100.0%; Pred. No. 0.0015;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 HWSYGLRP 9

DB 2 HWSYGLRP 9

RESULT 2

RHSHG

gonadoliberin - sheep

C:Species: Ovis orientalis aries, Ovis ammon aries (domestic sheep)

C>Date: 31-Dec-1991 #sequence_revision 31-Dec-1991 #text_change 18-Mar-1997

C:Accession: A93780; A01411

R:Burgus, R.; Butcher, M.; Amoss, M.; Ling, N.; Monahan, M.; Rivier, J.; Fellows, R.; F

Proc. Natl. Acad. Sci. U.S.A. 69, 278-282, 1972

A>Title: Primary structure of the ovine hypothalamic luteinizing hormone-releasing fact

A:Reference number: A93780; MUID:72094314; PMID:4550508

A:Accession: A93780

A:Molecule type: protein

hypothetical prote
hypothetical prote
protein C13A10.3 [
hypothetical prote
hypothetical prote
stage V sporulatio
conserved hypothet
peptidase, M16 fam
hypothetical prote
inorganic pyrophos
hypothetical prote
aldose reductase h
ylil protein precu
probable dehydroge
probable oxidoredu

30 37 68.5 486 2 T26483
31 37 68.5 812 2 T01618
32 36 66.7 315 2 A88043
33 36 66.7 327 2 E95323
34 36 66.7 368 2 T27432
35 36 66.7 460 2 D96971
36 36 66.7 517 2 F83573
37 36 66.7 927 2 H87568
38 35 64.8 82 2 G89446
39 35 64.8 300 2 T49933
40 35 64.8 308 2 D64696
41 35 64.8 316 2 A53440
42 35 64.8 371 1 E64821
43 35 64.8 371 2 E90743
44 35 64.8 371 2 A85594
45 35 64.8 373 2 AB0604

;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: Larson and Taylor
;; STREET: 727 Twenty-Third Street, South
;; CITY: Arlington
;; STATE: Virginia
;; COUNTRY: USA
;; ZIP: 22202
;;
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: Patentin Release #1.0, Version #1.25
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/343,883
;; FILING DATE: 17-NOV-1994
;; CLASSIFICATION: 424
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 07/946,495
;; FILING DATE: 09-NOV-1992
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: FR 9102513
;; FILING DATE: 01-MAR-1991
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: FR 9115289
;; FILING DATE: 10-DEC-1991
;; INFORMATION FOR SEQ ID NO: 1:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 10 amino acids
;; TYPE: amino acid
;; TOPOLOGY: linear
;; MOLECULE TYPE: Peptide
;; FEATURE:
;; NAME/KEY: Peptide
;; LOCATION: 10
;; OTHER INFORMATION: /label= NH2
;; OTHER INFORMATION: /note= "amidated glycine"
;;
;; FEATURE:
;; NAME/KEY: Peptide
;; LOCATION: 1
;; OTHER INFORMATION: /label= pyro
;; OTHER INFORMATION: /note= "pyroglutamic acid"
;;
;; PUBLICATION INFORMATION:
;; AUTHORS: Matsuo, H.
;; AUTHORS: Baba, Y.
;; AUTHORS: G. Naik, R. M.
;; AUTHORS: Arimura, A. V.
;; AUTHORS: Schally, A. V.
;; TITLE: Structure of the porcine LH- and
;; TITLE: FSH-releasing hormone. I. The proposed amino acid
;; TITLE: sequence.
;; JOURNAL: Biochem. Biophys. Res. Commun.
;; VOLUME: 43
;; ISSUE: 6
;; PAGES: 1334-1339
;; DATE: 1971
;; RELEVANT RESIDUES IN SEQ ID NO: 1: FROM 1 TO 10
US-08-343-883-1

Query Match 96.3%; Score 52; DB 1; Length 10;
Best Local Similarity 100.0%; Pred. NO. 0.0025;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Caps 0;

QY 2 HWSYGLRP 9
|||
Db 2 HWSYGLRP 9

Search completed: March 10, 2004, 09:28:50
Job time : 4.8716 secs


```
; NAME/KEY: Modified-site
; LOCATION: 1
; OTHER INFORMATION: /note= "Xaa at position 1 is a
; OTHER INFORMATION: 5-oxo-prolyl aminoacyl residue."
US-08-103-022-1

Query Match          96.3%; Score 52; DB 1; Length 10;
Best Local Similarity 100.0%; Pred. No. 0.0025;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 HWSYGLRP 9
   |||||
Db 2 HWSYGLRP 9

RESULT 13
US-07-897-680-1
; Sequence 1, Application US/07897680
; Patent No. 5446025
; GENERAL INFORMATION:
; APPLICANT: Fu Lu, Mou-Ying
; APPLICANT: Subba Rao, Gowdahalain N.
; APPLICANT: Lee, Dennis Y.
; TITLE OF INVENTION: Formulations and Method for the
; TITLE OF INVENTION: Percutaneous Administration of Leuprolide
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dept. 377 Abbott Laboratories
; STREET: One Abbott Park Road
; CITY: Abbott Park
; STATE: IL
; COUNTRY: USA
; ZIP: 60064-3500
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/897,680
; FILING DATE: 19920612
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Janssen, Jerry F.
; REGISTRATION NUMBER: 29,175
; REFERENCE/DOCKET NUMBER: 5165.US.01
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (708) 938-7742
; TELEFAX: (708) 937-9556
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 10 amino acids
; TYPE: AMINO ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 1
; OTHER INFORMATION: /note= "Xaa at position 1 is a
; OTHER INFORMATION: Pyro-glutamyl residue"
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 10
; OTHER INFORMATION: /note= "Xaa at position 10 is a
; OTHER INFORMATION: glycyl-amide residue"
US-07-897-680-1

Query Match          96.3%; Score 52; DB 1; Length 10;
Best Local Similarity 100.0%; Pred. No. 0.0025;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 HWSYGLRP 9
```

```
Db 2 HWSYGLRP 9
   |||||
RESULT 14
US-08-184-935-6
; Sequence 6, Application US/08184935
; Patent No. 5476770
; GENERAL INFORMATION:
; APPLICANT: PRADELLES, PHILIPPE
; TITLE OF INVENTION: IMMUNOMETRIC DETERMINATION OF AN ANTIGEN
; TITLE OF INVENTION: OR HAPTEN
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OBLON, SPIVAK, MCLELLAND, MAIER & NEUSTADT,
; ADDRESSEE: P.C.
; STREET: 1755 S. Jefferson Davis Highway, Suite 400
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/184,935
; FILING DATE: 24-JAN-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Oblon, No. 5476770man F.
; REGISTRATION NUMBER: 24,618
; REFERENCE/DOCKET NUMBER: 846-286-0
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 413-3000
; TELEFAX: (703) 413-2220
; TELEX: 248855 OPAT UR
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 10 amino acids
; TYPE: amino acid
; TOPOLOGY: unknown
; MOLECULE TYPE: peptide
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 10
; OTHER INFORMATION: /note= "C-terminal amide"
US-08-184-935-6

Query Match          96.3%; Score 52; DB 1; Length 10;
Best Local Similarity 100.0%; Pred. No. 0.0025;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 HWSYGLRP 9
   |||||
Db 2 HWSYGLRP 9

RESULT 15
US-08-343-883-1
; Sequence 1, Application US/08343883
; Patent No. 5573767
; GENERAL INFORMATION:
; APPLICANT: Dufour, Raymond J.
; APPLICANT: Roulet, Claude J.M.
; APPLICANT: Chouvet, Claire D.
; APPLICANT: Bonneau, Michel B.
; TITLE OF INVENTION: Method for improving the organoleptic
; TITLE OF INVENTION: qualities of the meat from uncastrated male domestic
; TITLE OF INVENTION: animals, vaccines which are usable in this method, new
; TITLE OF INVENTION: peptide, in particular for producing these vaccines...
; NUMBER OF SEQUENCES: 2
```

```
US-07-690-983D-32
; Sequence 32, Application US/07690983D
; Patent No. 5403586
; GENERAL INFORMATION:
; APPLICANT: RUSSELL-JONES, Gregory J.
; APPLICANT: STEWART, Andrew G.
; APPLICANT: TSONIS, Con G.
; TITLE OF INVENTION: FUSION PROTEINS
; NUMBER OF SEQUENCES: 47
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W.
; CITY: Washington, D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/690,983D
; FILING DATE: 25-JUN-1991
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/AU90/00373
; FILING DATE: 24-AUG-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 16786/148 CHAC
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)672-5300
; TELEFAX: (202)672-5399
; INFORMATION FOR SEQ ID NO: 32:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 10 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-07-690-983D-37
; Query Match 96.3%; Score 52; DB 1; Length 10;
; Best Local Similarity 100.0%; Pred. No. 0.0025;
; Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      2 HWSYGLRP 9
Db      2 HWSYGLRP 9

RESULT 11
US-07-690-983D-37
; Sequence 37, Application US/07690983D
; Patent No. 5403586
; GENERAL INFORMATION:
; APPLICANT: RUSSELL-JONES, Gregory J.
; APPLICANT: STEWART, Andrew G.
; APPLICANT: TSONIS, Con G.
; TITLE OF INVENTION: FUSION PROTEINS
; NUMBER OF SEQUENCES: 47
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W.
; CITY: Washington, D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/690,983D
; FILING DATE: 05-OCT-1993
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Janssen, Jerry F.
; REGISTRATION NUMBER: 29,175
; REFERENCE/DOCKET NUMBER: 5389.US.01
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (708) 938-7742
; TELEFAX: (708) 938-2623
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 10 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; FEATURE:

US-07-690-983D-37
; Query Match 96.3%; Score 52; DB 1; Length 10;
; Best Local Similarity 100.0%; Pred. No. 0.0025;
; Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      2 HWSYGLRP 9
Db      2 HWSYGLRP 9

RESULT 12
US-08-103-022-1
; Sequence 1, Application US/08103022
; Patent No. 5413990
; GENERAL INFORMATION:
; APPLICANT: Haviv, Fortuna
; APPLICANT: Fitzpatrick, Timothy D.
; APPLICANT: Swenson, Rolf E.
; APPLICANT: Nichols, Charles J.
; APPLICANT: Mott, Nicholas A.
; TITLE OF INVENTION: N-Terminus Modified Analogs of LHRH
; NUMBER OF SEQUENCES: 1
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Edward H. Gorman, Jr., Dept. 377
; STREET: Abbott Laboratories, One Abbott Park Road
; CITY: No. 5413990th Chicago
; STATE: IL
; COUNTRY: USA
; ZIP: 60064-3500
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/103,022
; FILING DATE: 05-OCT-1993
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Janssen, Jerry F.
; REGISTRATION NUMBER: 29,175
; REFERENCE/DOCKET NUMBER: 5389.US.01
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (708) 938-7742
; TELEFAX: (708) 938-2623
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 10 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; FEATURE:
```

RESULT 10

Query Match 96.3%; Score 52; DB 1; Length 10;
Best Local Similarity 100.0%; Pred. No. 0.0025;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 HWSYGLRP 9
|||||
Db 2 HWSYGLRP 9

RESULT 5

US-07-690-983D-1
; Sequence 1, Application US/07690983D
; Patent No. 5403586
; GENERAL INFORMATION:
; APPLICANT: RUSSELL-JONES, Gregory J.
; APPLICANT: STEWART, Andrew G.
; APPLICANT: TSONIS, Con G.
; TITLE OF INVENTION: FUSION PROTEINS
; NUMBER OF SEQUENCES: 47
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W.
; CITY: Washington, D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION NUMBER: US/07/690,983D
; FILING DATE: 25-JUN-1991
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/AU90/00373
; FILING DATE: 24-AUG-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 16786/148 CHAC
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)672-5300
; TELEFAX: (202)672-5399
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 10 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 1
; OTHER INFORMATION: /note= "Represents pyroglutamic
; OTHER INFORMATION: acid"
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 10
; OTHER INFORMATION: /note= "Represents glycineamide"
US-07-690-983D-1

Query Match 96.3%; Score 52; DB 1; Length 10;
Best Local Similarity 100.0%; Pred. No. 0.0025;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 HWSYGLRP 9
|||||
Db 2 HWSYGLRP 9

RESULT 6

US-07-690-983D-2

; Sequence 2, Application US/07690983D
; Patent No. 5403586
; GENERAL INFORMATION:
; APPLICANT: RUSSELL-JONES, Gregory J.
; APPLICANT: STEWART, Andrew G.
; APPLICANT: TSONIS, Con G.
; TITLE OF INVENTION: FUSION PROTEINS
; NUMBER OF SEQUENCES: 47
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W.
; CITY: Washington, D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION NUMBER: US/07/690,983D
; FILING DATE: 25-JUN-1991
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/AU90/00373
; FILING DATE: 24-AUG-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 16786/148 CHAC
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)672-5300
; TELEFAX: (202)672-5399
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 10 amino acids
; TYPE: amino acid
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
US-07-690-983D-2

Query Match 96.3%; Score 52; DB 1; Length 10;
Best Local Similarity 100.0%; Pred. No. 0.0025;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 HWSYGLRP 9
|||||
Db 2 HWSYGLRP 9

RESULT 7

US-07-690-983D-6
; Sequence 6, Application US/07690983D
; Patent No. 5403586
; GENERAL INFORMATION:
; APPLICANT: RUSSELL-JONES, Gregory J.
; APPLICANT: STEWART, Andrew G.
; APPLICANT: TSONIS, Con G.
; TITLE OF INVENTION: FUSION PROTEINS
; NUMBER OF SEQUENCES: 47
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W.
; CITY: Washington, D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/690,983D

OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/572,339
FILING DATE: 18-May-2000
CLASSIFICATION: <unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/474,555
FILING DATE: 07-JUN-1995
ATTORNEY/AGENT INFORMATION:
NAME: Saxe, Bernhard D.
REGISTRATION NUMBER: 28,665
REFERENCE/DOCKET NUMBER: 18733/584/IMIN
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202)672-5300
TELEFAX: (202)672-5399
TELEX: 904136
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 10 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
FEATURE:
NAME/KEY: Modified-site
LOCATION: 1
OTHER INFORMATION: /product= "OTHER"
/note= "The Xaa at position 1 = pyrroglutamic acid."
SEQUENCE DESCRIPTION: SEQ ID NO: 1:

US-09-572-339-1

Query Match 96.3%; Score 52; DB 1; Length 10;
Best Local Similarity 100.0%; Pred. No. 0.0025;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 HWSYGLRP 9
Db 2 HWSYGLRP 9

RESULT 3
US-07-714-540-9
Sequence 9, Application US/07714540
Patent No. 5282521
GENERAL INFORMATION:
APPLICANT: Almquist, Ronald G.
APPLICANT: Toll, Lawrence
TITLE OF INVENTION: ISOLATED ATRIAL PEPTIDE-DEGRADING
TITLE OF INVENTION: ENZYME AND NOVEL COMPOUNDS USEFUL AS INHIBITORS THEREOF
NUMBER OF SEQUENCES: 13
CORRESPONDENCE ADDRESS:
ADDRESSEE: Irell & Manella
STREET: 545 Middlefield Road, Suite 200
CITY: Menlo Park
STATE: California
COUNTRY: USA
ZIP: 94025
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/714,540
FILING DATE: 19910607
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: Reed, Diane E.
REGISTRATION NUMBER: 31,292
REFERENCE/DOCKET NUMBER: 8500-0135.00
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-327-7250
TELEFAX: 415-327-2951

TELEX: 706141
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 10 amino acids
TYPE: AMINO ACID
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-07-714-540-9

Query Match 96.3%; Score 52; DB 1; Length 10;
Best Local Similarity 100.0%; Pred. No. 0.0025;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 HWSYGLRP 9
Db 2 HWSYGLRP 9

RESULT 4
US-07-983-111-1
Sequence 1, Application US/07983111
Patent No. 5284657
GENERAL INFORMATION:
APPLICANT: Fu Lu, Mou-Ying
APPLICANT: Reiland, Thomas L.
TITLE OF INVENTION: "Compositions and Methods for the
TITLE OF INVENTION: Sublingual or Buccal Administration of Therapeutic Agents"
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Edward H. Gorman, Jr., Dept. 377-AP6D
STREET: One Abbott Park Road
CITY: Abbott Park
STATE: Illinois
COUNTRY: United States of America
ZIP: 60064-3500
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/983,111
FILING DATE: 30-NOV-1992
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/750,843
FILING DATE: 26-AUG-1991
ATTORNEY/AGENT INFORMATION:
NAME: Janssen, Jerry F.
REGISTRATION NUMBER: 29,175
REFERENCE/DOCKET NUMBER: 4848.US.01
TELECOMMUNICATION INFORMATION:
TELEPHONE: (708) 938-7742
TELEFAX: (708) 937-9556
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 10 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
FEATURE:
NAME/KEY: Modified-site
LOCATION: 1
OTHER INFORMATION: /note= "Xaa at position 1 is
OTHER INFORMATION: 5-oxo-proline"
FEATURE:
NAME/KEY: Modified-site
LOCATION: 10
OTHER INFORMATION: /note= "Xaa at position 10 is
OTHER INFORMATION: glycineamide"
US-07-983-111-1

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 10, 2004, 08:58:54 ; Search time 3.8716 Seconds
(without alignments)
133.345 Million cell updates/sec

Title: US-09-848-834A-1
Perfect score: 54
Sequence: 1 HWSYGLRPP 10

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA:
1: /cgn2_6/ptodata/2/aa/5A_COMB.pep.*
2: /cgn2_6/ptodata/2/aa/5B_COMB.pep.*
3: /cgn2_6/ptodata/2/aa/6A_COMB.pep.*
4: /cgn2_6/ptodata/2/aa/6B_COMB.pep.*
5: /cgn2_6/ptodata/2/aa/PTCUS_COMB.pep.*
6: /cgn2_6/ptodata/2/aa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	52	96.3	9	US-09-890-650-29	Sequence 29, Appl
2	52	96.3	10	US-09-572-339-1	Sequence 1, Appl
3	52	96.3	10	US-07-714-540-9	Sequence 9, Appl
4	52	96.3	10	US-07-983-111-1	Sequence 1, Appl
5	52	96.3	10	US-07-690-983D-1	Sequence 1, Appl
6	52	96.3	10	US-07-690-983D-2	Sequence 2, Appl
7	52	96.3	10	US-07-690-983D-6	Sequence 6, Appl
8	52	96.3	10	US-07-690-983D-7	Sequence 7, Appl
9	52	96.3	10	US-07-690-983D-8	Sequence 8, Appl
10	52	96.3	10	US-07-690-983D-32	Sequence 32, Appl
11	52	96.3	10	US-07-690-983D-37	Sequence 37, Appl
12	52	96.3	10	US-08-103-022-1	Sequence 1, Appl
13	52	96.3	10	US-07-897-680-1	Sequence 6, Appl
14	52	96.3	10	US-08-184-935-6	Sequence 5, Appl
15	52	96.3	10	US-08-343-883-1	Sequence 1, Appl
16	52	96.3	10	US-08-000-931-5	Sequence 5, Appl
17	52	96.3	10	US-08-428-488-22	Sequence 22, Appl
18	52	96.3	10	US-08-341-219-11	Sequence 11, Appl
19	52	96.3	10	US-08-453-588-2	Sequence 2, Appl
20	52	96.3	10	US-08-453-588-4	Sequence 4, Appl
21	52	96.3	10	US-08-453-588-6	Sequence 6, Appl
22	52	96.3	10	US-08-453-588-8	Sequence 8, Appl
23	52	96.3	10	US-08-453-588-10	Sequence 10, Appl
24	52	96.3	10	US-08-453-588-12	Sequence 12, Appl
25	52	96.3	10	US-08-453-588-14	Sequence 14, Appl
26	52	96.3	10	US-08-453-588-16	Sequence 16, Appl
27	52	96.3	10	US-08-453-588-19	Sequence 19, Appl

28	52	96.3	10	US-08-453-588-22	Sequence 22, Appl
29	52	96.3	10	US-08-188-223-1	Sequence 1, Appl
30	52	96.3	10	US-08-188-223-3	Sequence 3, Appl
31	52	96.3	10	US-08-188-223-8	Sequence 8, Appl
32	52	96.3	10	US-08-406-935-5	Sequence 5, Appl
33	52	96.3	10	US-08-591-917-1	Sequence 1, Appl
34	52	96.3	10	US-08-387-156-2	Sequence 2, Appl
35	52	96.3	10	US-08-474-555-1	Sequence 1, Appl
36	52	96.3	10	US-08-446-692-1	Sequence 1, Appl
37	52	96.3	10	US-08-243-678D-1	Sequence 1, Appl
38	52	96.3	10	US-08-796-598-6	Sequence 6, Appl
39	52	96.3	10	US-08-694-865-2	Sequence 2, Appl
40	52	96.3	10	US-08-694-865-18	Sequence 18, Appl
41	52	96.3	10	US-08-488-351A-1	Sequence 1, Appl
42	52	96.3	10	US-08-480-494B-1	Sequence 1, Appl
43	52	96.3	10	US-08-447-173A-6	Sequence 6, Appl
44	52	96.3	10	US-08-747-137-116	Sequence 116, Appl
45	52	96.3	10	US-08-878-748-2	Sequence 2, Appl

ALIGNMENTS

RESULT 1
US-09-890-650-29
; Sequence 29, Application US/09890650
; Patent No. 6685947
; GENERAL INFORMATION:
; APPLICANT: JACKSON, DAVID CHARLES
; APPLICANT: SAKRAVI, GHOSH
; APPLICANT: WALKER, JOHN
; TITLE OF INVENTION: T HELPER CELL EPITOPES
; FILE REFERENCE: 47-152
; CURRENT APPLICATION NUMBER: US/09/890,650
; CURRENT FILING DATE: 2001-11-05
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 29
; LENGTH: 9
; TYPE: PRT
; ORGANISM: Canis sp.
US-09-890-850-29

Query Match 96.3%; Score 52; DB 4; Length 9;
Best Local Similarity 100.0%; Pred. No. 3e+05;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 HWSYGLRP 9
Db 1 HWSYGLRP 8

RESULT 2
US-09-572-339-1
; Sequence 1, Application US/09572339
; Patent No. RE37710
; GENERAL INFORMATION:
; APPLICANT: MCBRIDE, William J.
; APPLICANT: KARACAY, Habibe
; APPLICANT: GRIFFITHS, Gary L.
; TITLE OF INVENTION: RADIOMETAL-BINDING ANALOGUES OF
LEUTENIZING HORMONE RELEASING HORMONE
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 3000 K Street, N.W., Suite 500
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20007-5109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible

```
; SEQ ID NO 3
; LENGTH: 10
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Gonadotrophin
; OTHER INFORMATION: Release Hormone Analogue
; NAME/KEY: MOD_RES
; LOCATION: (1)
; OTHER INFORMATION: Xaa at position 1 is PyroGlu; PYRROLIDONE
; OTHER INFORMATION: CARBOXYLIC ACID
; NAME/KEY: MOD_RES
; LOCATION: (10)
; OTHER INFORMATION: Xaa at position 10 is ethylamide;;
; PUBLICATION INFORMATION:
; PATENT DOCUMENT NUMBER: 5631226
; PATENT FILING DATE: 1993-07-07
; PUBLICATION DATE: 1997-05-20
; PUBLICATION INFORMATION:
; AUTHORS: Rousselle, Christophe
; JOURNAL: Mol. Pharmacol.
; VOLUME: 57
; PAGES: 679-686
; DATE: 2000
;
US-09-810-601-3

Query Match          96.3%; Score 52; DB 9; Length 10;
Best Local Similarity 100.0%; Pred. No. 0.022;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      2 HWSYGLRP 9
Db      2 HWSYGLRP 9

RESULT 6
US-09-810-601-4
; Sequence 4, Application US/09810601
; Patent No. US20020177545A1
; GENERAL INFORMATION:
; APPLICANT: Donovan, Stephen
; TITLE OF INVENTION: Compositions and Methods for Treating Gonadotrophin
; TITLE OF INVENTION: Related Illnesses
; FILE REFERENCE: 2947
; CURRENT APPLICATION NUMBER: US/09/810,601
; CURRENT FILING DATE: 2001-03-15
; NUMBER OF SEQ ID NOS: 44
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 10
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Gonadotrophin
; OTHER INFORMATION: Release Hormone Analogue
; NAME/KEY: MOD_RES
; LOCATION: (1)
; OTHER INFORMATION: Xaa at position 1 is PyroGlu; PYRROLIDONE
; OTHER INFORMATION: CARBOXYLIC ACID
; NAME/KEY: MOD_RES
; LOCATION: (10)
; OTHER INFORMATION: Xaa at position 10 is Aza-Gly-NH2;
; PUBLICATION INFORMATION:
; PATENT DOCUMENT NUMBER: 5631226
; PATENT FILING DATE: 1993-07-07
; PUBLICATION DATE: 1997-05-20
; PUBLICATION INFORMATION:
; AUTHORS: Rousselle, Christophe
; JOURNAL: Mol. Pharmacol.
; VOLUME: 57
; PAGES: 679-686
; DATE: 2000
;
US-09-810-601-4

Query Match          96.3%; Score 52; DB 9; Length 10;
Best Local Similarity 100.0%; Pred. No. 0.022;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      2 HWSYGLRP 9
Db      2 HWSYGLRP 9

RESULT 7
US-09-964-201A-28
; Sequence 28, Application US/09964201A
; Publication No. US20030091575A1
; GENERAL INFORMATION:
; APPLICANT: Kenten, John H
; APPLICANT: Tramontano, Alfonso
; APPLICANT: Pilon, Aprilie L
; APPLICANT: Lohnas, Gerald L
; APPLICANT: Roberts, Steven F
; TITLE OF INVENTION: HEAT-SHOCK FUSION-BASED VACCINE SYSTEM
; FILE REFERENCE: U.S. Patent Application No. US20030091575A1 09\026,276
; CURRENT APPLICATION NUMBER: US/09/964,201A
; CURRENT FILING DATE: 2002-05-21
; NUMBER OF SEQ ID NOS: 35
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 28
; LENGTH: 10
; TYPE: PRT
; ORGANISM: Porcine
;
US-09-964-201A-28

Query Match          96.3%; Score 52; DB 10; Length 10;
Best Local Similarity 100.0%; Pred. No. 0.022;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      2 HWSYGLRP 9
Db      2 HWSYGLRP 9

RESULT 8
US-09-964-201A-32
; Sequence 32, Application US/09964201A
; Publication No. US20030091575A1
; GENERAL INFORMATION:
; APPLICANT: Kenten, John H
; APPLICANT: Tramontano, Alfonso
; APPLICANT: Pilon, Aprilie L
; APPLICANT: Lohnas, Gerald L
; APPLICANT: Roberts, Steven F
; TITLE OF INVENTION: HEAT-SHOCK FUSION-BASED VACCINE SYSTEM
; FILE REFERENCE: U.S. Patent Application No. US20030091575A1 09\026,276
; CURRENT APPLICATION NUMBER: US/09/964,201A
; CURRENT FILING DATE: 2002-05-21
; NUMBER OF SEQ ID NOS: 35
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 32
; LENGTH: 10
; TYPE: PRT
; ORGANISM: Porcine
;
US-09-964-201A-32

Query Match          96.3%; Score 52; DB 10; Length 10;
Best Local Similarity 100.0%; Pred. No. 0.022;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      2 HWSYGLRP 9
Db      2 HWSYGLRP 9

RESULT 9
US-09-964-201A-32
; Sequence 32, Application US/09964201A
; Publication No. US20030091575A1
; GENERAL INFORMATION:
; APPLICANT: Kenten, John H
; APPLICANT: Tramontano, Alfonso
; APPLICANT: Pilon, Aprilie L
; APPLICANT: Lohnas, Gerald L
; APPLICANT: Roberts, Steven F
; TITLE OF INVENTION: HEAT-SHOCK FUSION-BASED VACCINE SYSTEM
; FILE REFERENCE: U.S. Patent Application No. US20030091575A1 09\026,276
; CURRENT APPLICATION NUMBER: US/09/964,201A
; CURRENT FILING DATE: 2002-05-21
; NUMBER OF SEQ ID NOS: 35
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 32
; LENGTH: 10
; TYPE: PRT
; ORGANISM: Porcine
;
US-09-964-201A-32

Query Match          96.3%; Score 52; DB 10; Length 10;
Best Local Similarity 100.0%; Pred. No. 0.022;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      2 HWSYGLRP 9
Db      2 HWSYGLRP 9

RESULT 9
```

```

US-09-305-924-1
; Sequence 1, Application US/09305924A
; Publication No. US20030091579A1
; GENERAL INFORMATION:
; APPLICANT: Jack G. Manns
; APPLICANT: Stephen D. Acres
; APPLICANT: Richard Harland
; TITLE OF INVENTION: METHODS OF RAISING ANIMALS FOR MEAT PRODUCTION
; FILE REFERENCE: 9001-0048
; CURRENT APPLICATION NUMBER: US/09/305,924A
; CURRENT FILING DATE: 1999-05-05
; EARLIER APPLICATION NUMBER: US 60/084,217
; EARLIER FILING DATE: 1998-05-05
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 10
; TYPE: PRT
; ORGANISM: GRH
; FEATURE:
; NAME/KEY: MOD_RES
; LOCATION: (1)_RES
; OTHER INFORMATION: Xaa is pyroglutamic acid
US-09-305-924-1
Query Match          96.3%; Score 52; DB 10; Length 10;
Best Local Similarity 100.0%; Pred. No. 0.022;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      2 HWSYGLRP 9
      |||||
Db      2 HWSYGLRP 9

RESULT 10
US-09-305-924-9
; Sequence 9, Application US/09305924A
; Publication No. US20030091579A1
; GENERAL INFORMATION:
; APPLICANT: Jack G. Manns
; APPLICANT: Stephen D. Acres
; APPLICANT: Richard Harland
; TITLE OF INVENTION: METHODS OF RAISING ANIMALS FOR MEAT PRODUCTION
; FILE REFERENCE: 9001-0048
; CURRENT APPLICATION NUMBER: US/09/305,924A
; CURRENT FILING DATE: 1999-05-05
; EARLIER APPLICATION NUMBER: US 60/084,217
; EARLIER FILING DATE: 1998-05-05
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 9
; LENGTH: 10
; TYPE: PRT
; ORGANISM: GRH
US-09-305-924-9
Query Match          96.3%; Score 52; DB 10; Length 10;
Best Local Similarity 100.0%; Pred. No. 0.022;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      2 HWSYGLRP 9
      |||||
Db      2 HWSYGLRP 9

RESULT 11
US-10-109-331-2
; Sequence 2, Application US/10109331
; Publication No. US20020165159A1
; GENERAL INFORMATION:
; APPLICANT: McCann, Samuel M.
; APPLICANT: Yu, Wen H.
; TITLE OF INVENTION: FSH-Releasing Peptides

```

```

; NUMBER OF SEQUENCES: 41
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: John H. Runnels
; STREET: P. O. Box 2471
; CITY: Baton Rouge
; STATE: LA
; COUNTRY: USA
; ZIP: 70821-2471
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25;
; WordPerfect 5.1; No. US20020165159A1eapad Version 4.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/109,331
; FILING DATE: 28-Mar-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/09/297,989
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Runnels, John H.
; REGISTRATION NUMBER: 33451
; REFERENCE/DOCKET NUMBER: 9703P-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (225) 387-3221
; TELEFAX: (225) 346-8049
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 10 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; FEATURE:
; OTHER INFORMATION: /note= "Xaa at 1 is pyro-Glu; Xaa
; at 10 is Gly-NH2"
; SEQUENCE DESCRIPTION: SEQ ID NO: 2:
US-10-109-331-2
Query Match          96.3%; Score 52; DB 13; Length 10;
Best Local Similarity 100.0%; Pred. No. 0.022;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      2 HWSYGLRP 9
      |||||
Db      2 HWSYGLRP 9

RESULT 12
US-10-184-126-1
; Sequence 1, Application US/10184126
; Publication No. US20020183257A1
; GENERAL INFORMATION:
; APPLICANT: EL TAYAR, Nabil
; APPLICANT: ZHAO, Xuan
; APPLICANT: BENTLEY, Michael D.
; TITLE OF INVENTION: PEG-LHRH ANALOG CONJUGATES
; FILE REFERENCE: EL-TAYAR=2A
; CURRENT APPLICATION NUMBER: US/10/184,126
; CURRENT FILING DATE: 2002-06-28
; PRIOR APPLICATION NUMBER: US/09/698,134
; PRIOR FILING DATE: 2000-10-30
; PRIOR APPLICATION NUMBER: 60/083,340
; PRIOR FILING DATE: 1998-04-28
; PRIOR APPLICATION NUMBER: PCT/US99/09160
; PRIOR FILING DATE: 1999-04-28
; NUMBER OF SEQ ID NOS: 1
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 10
; TYPE: PRT
; ORGANISM: Artificial Sequence

```